DOCUMENT RESUME

ED 272 883 CS 209 963

AUTHOR Swain, Murray G.
TITLE Effects of Stylistics on Essay Graders.
PUB DATE Jan 81
NOTE 241p.; M.Ed. Dissertation, University of Western Australia.
PUB TYPE Dissertations/Theses - Master Theses (042) -- Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Comparative Analysis; *Diction; Foreign Countries; *Grading; Secondary Education; *Writing Evaluation; *Writing Research
IDENTIFIERS Rater Reliability; *Stylistics; Voice (Rhetoric); Voice (Verbs)

ABSTRACT

Eight public examiners in English from the Western Australian Tertiary Admissions Examination marking panel and 103 final-year student teachers graded essays that were experimentally controlled for voice and diction. Control of these independent variables was achieved by providing four rewritten versions of the secondary student essays, each version identical to others except for one of four styles: active voice with Saxon diction, active voice with Latinate diction, passive voice with Saxon diction, and passive voice with Latinate diction. Grades were transformed to an interval scale using the Rasch Polychotomous Rating Model, and the transformed scores treated by repeated measures multivariate analysis of variance to test shifts in grader judgments from one version of an essay to the others. Results indicated a significant effect for diction, in favor of Latinate diction, when rated for overall quality, mechanics, logic, and organization. There was no effect for voice and no interaction effect. The students' original, unmodified essays received grades that placed them significantly higher than their Saxon versions but significantly lower than their Latinate versions. The experimental essay, with its intense uniformity of Latinate diction all in sentences of one voice, was preferred to the naturally written, variable, and flexible essay from the pen of a student. Among several implications arising from this study is the possibility that if English teachers instruct secondary students to write simply and clearly, students are being deceived, for such writing is not the most highly valued by teachers and public examiners alike. (Extensive appendixes include notes on the rewriting of essays and samples of stylistic versions, instructions to essay graders and examiners, and complete statistical results of data from three years.)
(Author/HTH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED272883

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EFFECTS OF STYLISTICS ON ESSAY GRADERS

A dissertation submitted in partial fulfilment
of the requirements for the degree of Master
of Education in the University
of Western Australia

by

Murray G. Swain, B.Ed. (Hons.)

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Murray G. Swain

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

January, 1981.

CS 209 963

ABSTRACT

Eight public examiners in English from the Western Australian Tertiary Admissions Examination marking panel and one hundred and three final-year student-teachers graded essays which were experimentally controlled for two elements that contribute to style, namely, voice and diction. Control of these independent variables was achieved by providing four re-written versions of secondary student essays, each version identical to others except for one of the four styles - Active voice with Saxon diction, Active voice with Latinate diction, Passive voice with Saxon diction and Passive voice with Latinate diction.

Grades were transformed to an interval scale using the Rasch Polychotomous Rating Model, and the transformed scores treated by repeated measures multivariate analysis of variance to test the dependent variable, that is, shifts in grader judgements from one version of an essay to the others.

Results showed significant effect for Diction such that essays with Latinate diction were rated significantly higher than the same essays in Saxon diction, and this effect occurred when graders rated for Overall Quality, Mechanics and Logic and Organisation. There was no Voice effect and no interaction

effect. The statistical outcomes were supported by written comments from the public examiners and show that graders overwhelmingly prefer a more cumbersome, indirect, abstract and verbose style rather than a simple, direct, clear and vigorous style in essays of the discussion or argument type. Graders make no distinction in grades between essays in the active voice or in the passive voice, and they regard an essay using Latinate diction as having fewer, or less important, errors in spelling, punctuation and usage than the same essay in Saxon diction. Further, they regard Latinate essays as being superior in logic and organisation of material.

The students' original, unmodified essays were found to receive grades that placed them significantly higher than their Saxon versions but significantly lower than their Latinate versions. The experimental essay, with its intense uniformity of Latinate diction all in sentences of one voice, was preferred to the naturally-written, variable and flexible essay from the pen of a student.

Amongst several implications arising from this study is one that raises the possibility that if English teachers instruct secondary students to write simply and clearly, then students are being deceived, for such writing is not the most highly valued by teachers and public examiners alike.

ACKNOWLEDGEMENTS

Grateful appreciation to Dr David Andrich of the University of Western Australia for his patient supervision.

Thanks are due also to Mr Brian Wolfenden, Senior Master, English, at Cyril Jackson Senior High School, for his assistance.

"He respects brevity, knowing its challenge, and attempts to cut away from complexity to clarity. His goal is simplicity - the flight of the seagull. He seeks the least complicated way of expressing the most complicated idea."

-- Donald M. Murray
A Writer Teaches Writing

TABLE OF CONTENTS

		Page
	ABSTRACT	ii
	ACKNOWLEDGEMENTS	iv
	LIST OF TABLES	viii
Chapter		
I	INTRODUCTION	1
	Background of the Problem	2
	Research Questions	4
II	REVIEW OF RELATED RESEARCH	7
	Review of Trends in English Curriculum and Teaching	7
	Research into Essay-Grading	19
III	THE STUDIES OF HAKE AND WILLIAMS IN ESSAY GRADING	31
IV	DESIGN OF THE STUDY	44
	Definition of Terms	45
	Hypotheses	48
	Subjects	50
	Procedures	53
	Statistical Treatment and Analysis	55
	Limitations of the Study	58
	Assumptions	60
V	RESULTS AND DISCUSSION	62
	Reliability of 1978 Graders	63
	1976 Results	65
	1977 Results	78
	1978 Results	83

Chapter		Page
VI	SUMMARY, CONCLUSIONS AND IMPLICATIONS	100
BIBLIOGRAPHY	. . . . .	107
APPENDIX A:	. . . . .	114
	Notes on the Re-Writing of Essays and Samples of Stylistic Versions	
APPENDIX B:	. . . . .	147
	Instructions to Essay Graders and Grades and Comments of Eight Examiners	
APPENDIX C:	. . . . .	173
	The Use of the Rasch Polychotomous Rating Model	
APPENDIX D:	. . . . .	181
	Complete Statistical Results of Analysis of Variance of 1976 Data	
APPENDIX E:	. . . . .	189
	Complete Statistical Results of Analysis of Variance of 1977 Data	
APPENDIX F:	. . . . .	197
	Complete Statistical Results of Analysis of Variance of 1978 Data	

LIST OF TABLES

Table		Page
I	Style Characteristics of Re-written Essays . . .	46
II	Chi-square Values for Observed and Expected Proportions of Grader Awards to Each Grade-Score on 1978 "Padding" Essays	64
III	Means of Transformed Scores for Each Essay on Each Grading Dimension: 1976 Data	66
IV	Cell Means of Transformed Scores for Essays by Style on Each Grading Dimension: 1976 Data . . .	67
V	Analysis of Variance Table for Interaction Effects of Voice and Diction (1976)	68
VI	Analysis of Variance Table for Main Effects of Voice (1976).	69
VII	Analysis of Variance Table for Main Effects of Diction (1976)	71
VIII	Means of Transformed Scores for Each Essay on Each Grading Dimension: 1977 Data	78
IX	Cell Means of Transformed Scores for Essays by Style on Each Grading Dimension: 1977 Data . . .	79
X	Analysis of Variance Table for Interaction effects of Voice and Diction (1977)	79
XI	Analysis of Variance Table for Main Effects of Voice (1977)	80

Table		Page
XII	Analysis of Variance Table for Main Effects of Diction (1977)	81
XIII	Means of Transformed Scores for Each Modified Essay on Each Grading Dimension: 1978 Data . .	84
XIV	Cell Means of Transformed Scores for Modified Essays by Style on Each Grading Dimension: 1978 Data	85
XV	Analysis of Variance Table for Interaction Effects of Voice and Diction (1978)	85
XVI	Analysis of Variance Table for Main Effects of Voice (1978)	86
XVII	Analysis of Variance Table for Main Effects of Diction (1978)	87
XVIII	Means of Transformed Scores for all Modified and all Original Essays on each Grading Dimension (1978)	91
XIX	Cell Means of Transformed Scores for Essays by Modified and Original Styles on Each Grading Dimension (1978)	92
XX	Means and Standard Deviations of Differences Between Modified and Original Essays by Style and Grading Dimension	93
XXI	Analysis of Variance Table for Comparison of Style I (AS) and Original Essays (1978).	94

Table		Page
XXIV	Analysis of Variance Table for Comparison of Style II (AL) and Original Essays (1978)	95
XXIII	Analysis of Variance Table for Comparison of Style III (PS) and Original Essays (1978) . . .	96
XXIV	Analysis of Variance Table for Comparison of Style IV (PL) and Original Essays (1978). . . .	97
	APPENDIX C TABLES:	
C I	Chi-Square Values for Observed and Expected Proportions of 1977 Grader Awards to Each Grade-Score for Essays on Overall Grading Dimension	178
C II	Chi-Square Values for Observed and Expected Proportions of 1977 Grader Awards to Each Grade-Score for Essays on Mechanics Grading Dimension	179
C III	Chi-Square Values for Observed and Expected Proportions of 1977 Grader Awards to Each Grade-Score for Essays on Logic Grading Dimension	180
	APPENDIX D TABLES:	
D I	Means, Variances and Standard Deviations of 1976 Variables	182
D II	Multivariate Analysis of Variance Summary Table for Dependent Variables PL-AL (1976) . . .	183

Table		Page
D III	Multivariate Analysis of Variance Summary Table for Dependent Variables AL-PS (1976) . . .	184
D IV	Multivariate Analysis of Variance Summary Table for Dependent Variables PS-AS (1976) . . .	185
D V	Multivariate Analysis of Variance Summary Table for Dependent Variables A-P (1976)	186
D VI	Multivariate Analysis of Variance Summary Table for Dependent Variables S-L (1976)	187
D VII	Multivariate Analysis of Variance Summary Table for Dependent Variables AP.SL (1976) . . .	188
	APPENDIX E TABLES:	
E I	Means, Variances and Standard Deviations of 1977 Variables	190
E II	Multivariate Analysis of Variance Summary Table for Dependent Variables PL-AL (1977) . . .	191
E III	Multivariate Analysis of Variance Summary Table for Dependent Variables AL-PS (1977) . . .	192
E IV	Multivariate Analysis of Variance Summary Table for Dependent Variables PS-AS (1977) . . .	193
E V	Multivariate Analysis of Variance Summary Table for Dependent Variables A-P (1977)	194

Table | Page

E VI Multivariate Analysis of Variance Summary Table for Dependent Variables S-L (1977) 195

E VII Multivariate Analysis of Variance Summary Table for Dependent Variables AP.SL (1977) . . . 196

APPENDIX F TABLES

F I Means, Variances and Standard Deviations of 1978 Variables 198

F II Multivariate Analysis of Variance Summary Table for Dependent Variables PL-AL (1978) . . . 199

F III Multivariate Analysis of Variance Summary Table for Dependent Variables AL-PS (1978) . . . 200

F IV Multivariate Analysis of Variance Summary Table for Dependent Variables PS-AS (1978) . . . 201

F V Multivariate Analysis of Variance Summary Table for Dependent Variables A-P (1978) 202

F VI Multivariate Analysis of Variance Summary Table for Dependent Variables S-L (1978) 203

F VII Multivariate Analysis of Variance Summary Table for Dependent Variables AP.SL (1978) . . . 204

CHAPTER I

INTRODUCTION

Each year in Australia, thousands of high school students write essays under examination conditions for their matriculation and school-leaving certificates. These essays are collected and brought to the assistant-examiners of English to grade. Then follows the long, and expensive, period of time during which the minds of graders confront the assorted writings of candidates. Hopefully, valid and reliable assessments are made of each candidate's writing ability.

Those moments when the grader assesses the content, relevance, style, logic, organisation and structure of the essay are crucial in at least three ways. They are crucial for the candidate, who hopes that whoever grades his paper does so fairly and accurately, without misinterpretation of his evidence of language ability. They are crucial for the examiner, who must be sensitive and responsive to the components of content and expression so that he awards a grade that validly represents the ability of the candidate. Third, these moments are the culmination of thousands of hours of instruction and development in language skills, crystallising in one act the thinking and work of teachers, academics, administrators and examiners. No matter how informed, skilful and dedicated the work of these people, and no matter how large the body of research and knowledge about English teaching, about children's language, and of measurement theory and practice, it is all lost if the grader's reactions are invalid. The

examiner represents the final test of what has gone on before. The work of teacher and students, the current philosophy and practices of English teaching, the findings of research in English curriculum, linguistics, psychology, sociology and education - all are brought to account in those crucial moments.

Of course, the examiner is not the only person who grades essays. The English teacher in the high school and technical school, the primary school teacher, and the secondary teacher in the subject areas other than English are all accountable to the same body of knowledge and practice in English curriculum and the body of research knowledge in measurement and evaluation, when they evaluate students' writing.

It is essential that writing ability be evaluated accurately. The present study investigates certain problems associated with the validity of essay-graders' judgements, particularly in relation to the current bodies of knowledge referred to above in English curriculum and measurement and evaluation.

Background of the Problem

The last decade has seen the emergence and acceptance of new principles in English curriculum, and, accordingly, revised practices in evaluating language ability in children. Until the mid-1960's, English teaching concerned itself with the transmission of a body of skills that had to be mastered by children, often in imitation of selected models from classical literature. English was seen to be "respectable" as a

discipline, with a body of theory and substantial knowledge which could be passed on to students in the same way as scientific knowledge, mathematics or history. Written expression was seen as a sequence of activities undertaken by learners so as to rehearse their mastery of desirable and clearly-defined skills. Correct grammar, sentence-structure, and rich vocabulary were taught as components of a "transmission model" of English curriculum and teaching.

With the 1966 Dartmouth Seminar in England came the acceptance of a different model of English curriculum and teaching. The language that children use in their own families and social circles was acknowledged as being important to the child. It became the school's task to carry on the development of a child's own use of language, acknowledging that the language he uses actually works for him at home and should not be discredited. The curriculum should enlarge his language activities and experiences at a level that the learner can manage, so that he learns to handle language that is flexible enough to meet the needs of his widening encounters with his world. English curriculum became a "growth" model, an interpretation and interaction model which involved the child in situations that required his mastery of communication skills as he was ready for them.

In this new context, written expression is now seen as an evolving and continuous development of the child's mastery of skills and thought in writing activities. The activities and tasks that the teacher provides are not based solely on models from literature to be imitated, but on the emerging needs and competencies that the child displays. He should

be writing about things that matter to him, for audiences that are interested in what he has to say, and for all sorts of purposes - social, formal, personal, poetic, public debate, academic, and so on. Certainly, a teacher will supply imitation models and activities, just as he will set topics and tasks, as teachers always have done, but the reason will be to stimulate the learner at his present stage of development so that he keeps moving ahead, not because there is an absolute body of knowledge that "ought" to be transmitted in regulated fashion to inert learners.

The evaluation of writing ability had to change to suit the new English curriculum, and such changes are described fully in Chapter II. In essence, one would expect to see teachers and public examiners in English rewarding those students who show versatility of expression and the ability to match content, style, purpose and expression with the occasion and audiences for which they are writing. It follows that no one kind of writing can be deemed to be the "right" one for all occasions.

Research Questions

With these new demands on the evaluation of writing ability, an important question is whether graders do, or can, meet these demands. Do they reward only certain kinds of writing, or do they acknowledge effective communication that has versatility of expression? Is there a legacy from the past philosophy of English curriculum whereby teachers and examiners still see the desirability of writing in one "correct" way? Maybe some teachers continue to reward such essays.

It is expected that teachers and examiners would reward appropriate and effective writing.

Appropriate writing uses a style of expression that suits purpose, audience and subject (Boulton, 1954). Sometimes writing should be formal and in weighty register; at other times it should not. But one principle has always been associated with good writing: that principle is to write clearly, directly to the point, as simply as possible, and without unnecessary complexity. All manuals, guides, texts and treatises on writing advocate clarity of meaning and expression in writing. To acknowledge all such reference sources here would be impossible, but well-known and accepted sources include Orwell (1946), Scott (1948), King (1955), Warner (1961), Boulton (1954) and Murray (1968). The recognised masters of the English essay, such as Bacon, Orwell, Hazlitt, A. Huxley, Russell and Chesterton, write with directness and clarity and as much vigour and simplicity as possible.

In the light of the above, the questions still remain to be asked: do teachers and examiners reward appropriate writing, or do they reward only a certain kind of writing? Do teachers support the all-important principle of clear and direct writing in their teaching and grading? Could there be some anomalies in what they reward and what they ought to reward according to accepted principles and practice? It could be the case that teachers espouse one set of principles of written expression when they are teaching, such as writing directly and simply, but when it comes to marking, they

could be tacitly observing a different set of principles and rewarding those children who do not write simply and directly. Absurd as this contradiction appears, there has been recent research conducted in the U.S.A. that suggests that this could, in fact, be true (Hake and Williams, 1975, 1979).

In order to seek answers to these questions, the present study took shape using style as the independent variable so that graders could be confronted with one essay re-written in various versions, as Hake and Williams had done. The style that the graders approved of might then be rewarded with higher grades.

The research questions and general approach of this study have been briefly outlined in this chapter to show how they were developed from the current background of English curriculum. They arise also from recent research in the evaluation of writing ability. Before details of the study can be explained further, it is necessary now to turn to the literature in both English curriculum and research into essay-grading, and to examine them in detail.

CHAPTER II

REVIEW OF RELATED RESEARCH

This study stands against two backgrounds. On the one hand, it is related to the current views of English curriculum and teaching; and on the other, it follows a new trend in design of experimental studies in essay-grading. To set the study in a complete context, each background will be described, first by reviewing the literature in English curriculum, and then by outlining the research into essay-grading.

Review of Trends in English Curriculum and Teaching

There have been significant changes in the philosophy of English teaching since the beginning of this century. As these changes occurred, one would have expected to see appropriate changes in methods and criteria of evaluation in English, for there should always be a logical relationship between curriculum goals, curriculum activities and resources, and curriculum evaluation. This logical connection should be apparent regardless of the prevailing curriculum and pedagogical principles of the time. It would be absurd to find, for example, that the criteria by which graders judged essay-writing ability were different from the criteria which had been taught to students by informed teachers practising contemporary theory of writing.

According to Dixon (1975), earlier English curricula have followed a skills model, where literary or grammatical skills have been listed and rehearsed by students, or they have followed the cultural heritage model, where great literature has been held up to learners as models for analysis and imitation. Devine (1968) identified trends in the American context by examining several reports and commission findings that had been published.

Devine turned to Eaton's College Requirements in English Entrance Examinations of 1894[1] to find that secondary school English was characterised by grammatical and rhetorical analysis of selected literary classics, involving parsing, exegesis, sentence diagramming, classification of figures of speech, critical commentaries and memorising literary history and biographies. "English was a discipline to be mastered by learners rather than a general name for a host of useful communication skills to enrich the lives of students" (Devine, 1968, p 19).

Modification of the analytic model came with the publication of the Hosic Report of 1917[2]. English was seen less as a body of facts to be learned and more as a set of competencies to be mastered through study and practice of basic communication skills. By 1935, an experimental curriculum was advocated by the Commission of the National Council of Teachers of English whose report[3]

---

[1] *Eaton, A. W. College Requirements in English Entrance Examinations. Boston: Ginn, 1894.*

[2] *Published as Hosic, James F. (compiler). Reorganisation of English in Secondary Schools. Washington: U.S. Bureau of Education, 1917.*

[3] *Curriculum Commission of the National Council of Teachers of English. An Experience Curriculum in English. New York: Appleton-Century-Crofts, 1935.*

recommended that teachers seek out "experience strands" through the grade levels, by which strands the diverse needs and backgrounds of students would be met. The curriculum was flexible and non-standardised so that it could cater for the development of language in students of all backgrounds and socio-cultural experiences across U.S.A.. It did not focus only on language, composition and literature; it aimed to extend students' competence in all communication skills over a wide range of social, vocational and literary purposes. Its composition requirements would have encouraged the development in learners of writing skills for varied purposes and situations. In this way, it was similar to the present-day English curriculum in Australia.

After the social changes effected by the World War, the National Council of Teachers of English realised that it needed to extend its work. In 1952, it published the first volume of The English Language Arts[4]. This report presented an overview of the thinking and current practices; it also described a method of approach to curriculum building. English was seen to be the processes of reading, writing, speaking and listening - a curriculum of activities which would lead students to acquire and develop language through use, not through the memorisation of content or the meticulously analytic study of classic writings. In many ways, this curriculum was identical to the current thinking in Australia and the United Kingdom in 1980. It was some thirty years ahead of practices in the U.K.

---

[4] *Commission on the English Curriculum of the National Council of Teachers of English. The English Language Arts. New York: Appleton-Century Crofts, 1952.*

Devine noted that the 1952 curriculum report failed to have lasting impact because America was shortly to launch itself into the international space-race and re-appraise its educational efforts. Accordingly, in 1959, the College Entrance Examination Board appointed its own Commission on English because it felt the need to define the diffuse concept of English curriculum (Devine, 1968). It published its report in 1965, entitled Freedom and Discipline in English[5]. It recommended that "the English program be defined as the study of language, literature and composition, with language as the core of the program"; anything not related to such language study was to be excluded (Devine, 1968, p 21).

The trend that the 1965 report began is current in U.S.A. today. Students study rhetoric and composition skills that are prescribed by boards of school districts and colleges. Following the work of Chomsky (1957, 1972) and other linguists, and supported by a flood of applied research in writing-instruction (Witte and Davis, 1980), American curricula in English at secondary and college level have focussed on analysis of writing skills involving transformational grammar, sentence-expansion and sentence-elaboration, in concentrated efforts to accelerate the syntactic maturity of students (Bennett, 1979). This energetic thrust in language-study, informed and research-based as it is, reflects the same assumptions about English as were current and generally approved up to the 1920's. Such a trend has been opposed by Moffett (1968, 1976) in U.S.A. itself, and by the emergent thinking of English researchers and writers on the other side of the Atlantic (Bennett, 1979). The assumption

---

[5] *Commission on English. Freedom and Discipline in English. New York: College Entrance Examination Board, 1965.*

in U.S.A. generally is that English is a discipline with a clearly-developed body of knowledge existing independent of the learner and which, if mastered, will enable the learner to cope with and understand the demands of life (Devine, 1968).

The current language-study model in America is diametrically opposed to the current "growth" model of English curriculum held in United Kingdom and Australia. The "growth" model concept is widely held in Australia (Schoenheimer, 1972), and arose from the outcomes of the famous Dartmouth Seminar on English of 1966. Dixon (1975) reviewed and interpreted the outcomes of the Seminar which held that English is "a broad field, embracing a range of competencies in reading, writing, speaking and listening", and, further, "a body of skills in communication in self-perception, and in self-understanding" (Devine, 1968, p 22).

This growth model involves a development of both language and personal growth in intellect, in social and moral concepts, in sensitivity, tolerance, self-reliance and other attributes characteristic of the highly expressive individual. The curriculum in English according to this model does not transmit a pre-determined sequence of skills to passive learners; it involves learners in interacting with and intepreting the world around them in what Bruner would call a "context-sensitive" curriculum (Bruner, 1972).

The large body of literature that has followed Dartmouth is based on the view that English is learned, not taught. The English

curriculum should give all its emphasis to language in actual use because language is the most significant way of symbolising and representing experience, of making sense of one's experience in the world (Kelly, 1963; Britton, 1970, 1977). The construction and reconstruction of experience is reflected in how a person construes his linguistic reality. In a sense even more intense than Dewey's experiential reality (Dewey, 1916), a person's linguistic reality is the reality by which he lives. For the reflection upon and the interpretation of an experience depend very largely upon the terms in which it has been symbolised. Children need an agent to help them develop symbolisation of experience into language just as much as they need the experience itself. The teacher is the agent who guides and models the learner in school (Dixon, 1975; Barnes et al, 1971), while the parent and family and peer group provide models out of school (Bernstein, 1971; Halliday, 1977a, 1977b). Absolute standards of correctness cannot be prescribed in the light of socio-linguistic research (Britton, 1970; Guth, 1973; Halliday, 1977a), and the child should be encouraged to use his own language in initial attempts to interpret surroundings and experiences (Dixon, 1975; Rosen, 1973; Labov, 1970; Lawton, 1966). With close attention to the needs and progress of the individual learner, the teacher supplies alternative models of expression, such as "standard English" and approved conventions, as the learner shows he is ready for them (Martin, 1971; Martin et al, 1976a, 1976b; Dixon, 1975; Burgess et al, 1973; Rosen, 1973; Barnes et al, 1971, 1975; Bruner, 1972; Creber, 1972).

It is clear that English curriculum teaching in U.K. and U.S.A. are at polarised extremes, and it would be naive to assume that either one has all the advantages and no disadvantages. Teachers, writers of textbooks and curriculum designers should not be trapped by what has been called "the faulty dilemma" (Devine, 1968); that is, some sensible merging of both sets of views is probably desirable, rather than advocating one extreme to the exclusion of the other.

It is impossible to say with certainty which end of the continuum is actually being favoured in practice by teachers in Western Australia at the present time. Informed teachers in this State, such as those who are members of the English Teachers' Association of Western Australia, a body closely affiliated with the Australian Association for the Teaching of English, are, no doubt, aware of the British views more than of the American views. This is because the journal of the Australian Association for the Teaching of English has clearly supported the "growth" model in its editions over the past ten years. From the writer's observations, Western Australian teachers are more likely to be practising according to the post-Dartmouth "growth" model. But it would be inaccurate to conclude from the above that this model predominates. There is probably a variety of philosophies in practice, and a large proportion of teachers who follow the skills model and/or the cultural heritage model.

Some evidence can be summoned. One indication of the goals of composition instruction in upper secondary schools is the list of readings suggested by the Tertiary Admissions Examination Board.

Both the reading lists for English and the writing-tasks themselves that are set in each year's examination encourage and invite a wide range of interests and responses from students. The topics for essays include wide-ranging material for content and a wide variety of models of expression such as argument, parody, satire, speeches, exposition, journal, magazine and newspaper articles, poetry, debate and discussion. Half of the writing requirements deals with analysis and comment upon unseen passages; the other half deals with writing on two topics selected freely from a list of twelve or thirteen. No compulsory essay topic has been set since the 1972 examination.

In lower secondary school, and in primary, students in Western Australia have more opportunity for expressive and poetic writing, loosely called "creative writing". Upper secondary students are directed more to writing that deals with content from reading and with topical social issues; students are expected to be more discursive and analytic (Bennett and Walker, 1978).

It is quite clear, then, that students in Western Australian secondary schools have the opportunities and the encouragement, if not direct instruction, to become flexible in their writing styles, to write for different purposes, to arrange their own material and approach to topics, and to write with different audiences in mind. Bennett and Walker (1978) supply evidence to substantiate this conclusion; so do the papers set by the Tertiary Admissions Examination body and so do the materials and teachers' guides

distributed by the Curriculum Branch of the Western Australian Education Department. A constant principle underlying the current philosophy of English teaching in Western Australia is that "good readers make good writers".

Versatility and flexibility of expression in students' writing is a characteristic that is to be expected in the light of the influence on British thinking of socio-linguistic research. Language has social functions. Part of the role-behaviour acted out by any person is the use of the appropriate register and code of communication (Bernstein, 1971; Halliday, 1977b). If a person is restricted in the communication skills which are expected in given social contexts, he will find himself being unable to fulfil the role satisfactorily; he will meet disapproval, and experience wider social-distance. The ability to use the appropriate diction, grammar, register and syntax, quite apart from conveying the semantic content considered appropriate for the occasion, is a large part of the capacity of an individual to take on the role at all (Bernstein, 1971; Labov, 1970; Lawton, 1966, 1968; Barnes et al, 1971).

Certain sorts of communication are considered appropriate, and are valued for social reasons, such as prestige and group-cohesiveness. It is not surprising, then, that role enlargement through drama and talk is an important part of both Dixon's growth model (Dixon, 1975) and Moffett's discourse model (Moffett, 1968, 1976). In Australia,

Robinson (1976) has recommended that children be taught to speak and write appropriately for any dominant sub-culture in their society; Eltis (1980) recommends that children should handle the socio-linguistic realities of our society with tolerance and understanding.

The evidence outlined so far indicates that Australian teachers, and certainly Western Australian teachers, are expected to support the growth model, particularly in terms of encouraging flexibility of expression and versatility in writing skills. Other views about composition teaching have been identified by Diamond (1979). He asked ninety-three Grade 10 teachers from twenty-two large state high schools in Brisbane and Ipswich to sort out eighty statements about English and written-expression instruction. The teachers, who were all volunteers, could sort their cards into as many categories as they each wished, ranging from "Agree" to "Disagree", and they were to comment upon their categories so chosen.

Using Latent Partition Analysis (Wiley, 1967), Diamond found that the most universally-perceived category of the total of twenty that were discerned in the responses was The Importance of Talk:

> "The belief it enshrines is that writing is so deeply dependent upon speech that lack of proficiency in talking results in minimum proficiency in writing as well."
>
> (Diamond, 1979, p 61)

He suggested that this category reflects the "growth" model.

Another of "the more widely-recognised" categories was, in fact, Growth:

> "It is argued that the chief concern of English is to enable the student to develop language so that he can explore his personal experience ... Contrary to the traditional view, the growth-centred belief is that the most important writing is of a poetic or imaginative nature, since it permits the young person to explore personal feelings and thoughts. This is current dogma ... (which) ... recalls progressive pedagogy ..."
>
> (Diamond, 1979, p 61)

Diamond found other clear groupings of responses, notably: Teacher-Student Relationship (trust and mutuality); The Mechanical Skills (traditional dispensing to children of mechanical skills); and Natural Development (young people naturally learn to write as they mature). It was clear that teachers believe in different, even competing, paradigms, as the highest probability of occurrence for any one of the categories was .58, and most were around .40. He pointed out that a question that needs to be answered is whether or not these ideas of teachers are actually being implemented in practice.

In summary, the following characteristics of the current "philosophy" of curriculum and teaching in English seem to apply in Western Australia, and probably over Australia as a whole:

(1) the "growth" model, with its emphasis on the social function of language and its role in personal development, is widely acknowledged

(2) the writing-tasks of students should involve them in much talking and discussion

(3) there is no evidence that shows widespread support for the skills model as prevails in the U.S.A.

(4) flexibility of expression (which would include style) is an important criterion in evaluating a student's writing ability, and so is versatility in employing modes of written communication (speeches, exposition, and so on)

(5) English is taught by providing stimulating experiences through reading and other media on topical and relevant content, and such activities promote ability in understanding and handling wider ranges of communication skills through imitation, rehearsal and instruction.

It is not likely, then, that only one kind of writing style is deemed to be the "right" one by composition teachers in Western Australia. One would expect to see essay-graders rewarding those essays that direct their style of expression appropriately towards the purpose of writing and towards the audience for whom it is meant.

From the foregoing review of trends in English curriculum, certain characteristics of essay-graders would be expected and approved today. But there are other characteristics of essay-graders that have been found through research into essay-grading.

Some of these characteristics are to be expected, some to be approved, and some to be changed if valid and reliable grading is to be done.

Research into Essay-Grading

Traditionally, research has focussed on ways of improving inter-grader reliability and on ways of making decisions about the quality of students' writing.

A thorough review by Gosling (1966) of the earliest studies of reliability in essay-grading described the work of Willing in 1918[6] and Hudelson, 1923[7]. Willing set up a rating scale to be used by markers who could also refer to model essays which exemplified values along the scale. Essays could be evaluated by matching them with the model essays at particular points on the scale. Willing calculated standard scores for certain points on the 9-point scale, but offered no data on reliability. Hudelson developed a similar rating scale and found inter-grader correlations to be in excess of 0.8. He attributed such a high correlation to the fact that only one topic had been set for the students to write on, and all raters had been trained before the experiment. The results of these early studies showed two factors that affect grader reliability - the need for a single, defined topic and the need to train graders.

---

[6] *Willing, M. H. Scale for Measuring Written Composition. Bloomington, Illinois: Public School Publishing Co., 1926.*

[7] *Hudelson, E. The Hudelson Typical Composition Ability Scale. Bloomington, Illinois: Public School Publishing Co., 1923.*

Focus shifted to other ways of using a rating scale, and Gosling (1966) reviewed how, in 1926, Thomson and Bailes[8] had primary school teachers grade young children's compositions using a general impression approach. The inter-correlations ranged from 0.67 to 0.82 for the six male head teachers, and from 0.52 to 0.73 for the five women teachers. Clearly, there had been greater variation in graders' awards because there was no guidance of markers through the supply of a set of model compositions. Gosling reported Stalnaker's 1934 study[9] which revealed much higher grader reliability when the essay topics were more closely defined for the students who were writing the essays.

Hartog and Rhodes (1936) studied marker reliability amongst examiners for a university scholarship in England. They found the inter-marker reliability coefficient to be only 0.51. Ten examiners marked 75 essays using a close, analytical approach and another 75 essays using general impression. The analytical procedure involved awarding marks to various elements of composition such as vocabulary, craftsmanship and accuracy. It was clear that analytic marking was no more precise or reliable than impression marking, since the variability amongst markers was so high, regardless of method. A similar study by Cast (1939) yielded comparable results when graders used four different methods. The highest average correlation coefficient was 0.6.

---

[8] *Thomson, G. and Bailes, S. The reliability of essay works. Forum of Education, 1926, 4, 85.*

[9] *Stalnaker, J. M. The validity of the University of Chicago Qualifying Examination. English Journal, 1934, 23, 384-388.*

_____. The construction and results of a 12 hour test in English composition. School and Society, 1934, 39, 218-224.

Later, Hartog chaired an investigation sub-committee of the International Institute Examinations Inquiry (Hartog, 1941). He had eight experienced examiners discuss those elements of composition that they wished to take into consideration in their marking of six hundred essays from secondary-school students. The "categories", as they were called, were Sense; Spelling; Punctuation (which included Formal Paragraphing); Grammar; Vocabulary; Sentence Structure; and General Impression, which had to be marked last. Although they regarded Sense as being a very important category, the examiners showed no "greater agreement when marking for a well-defined category, such as Sense, than when they mark(ed) for General Impression" (Hartog, 1941, p 128). In fact, the inter-examiner correlations ranged from -0.04 to 0.95.

Hartog's investigations went further. He advocated that an estimation of an essay's "true" score be calculated from the mean of all graders for each script on each category. Inter-category correlations were computed, and a factor analysis of these yielded a single factor - halo effect. Evidently, the examiners had been unable to "free their minds entirely from the General Impression" (p 129) when marking the other categories.

The literature had made it clear by the 1950's that reliability of essay-grading required more specifically-stated topics for compositions, eradication of halo effect, and a rating scale that did not depend on the marker's own particular interpretations for its application.

Other approaches were used in attempts to improve grader reliability, notably multiple-marking. Wiseman (1949) established the Devon System by which teams of four examiners for the 11+ English examination pooled their marks for each essay. Like Hartog, Wiseman maintained that the true score for essays is more closely approached the more markers there are who pool their marks. Wiseman's study over the years 1943 and 1948 was notable, too, because it was the first successful attempt to improve "system reliability" in a public examination. Furthermore, he anticipated the need to identify and act upon inconsistent markers, though the way he did it was somewhat different from the emergent practices today. He had each team re-mark a sample of papers three months after the examination. Markers used rapid general-impression on both occasions. A mark-remark correlation for each marker was calculated, and any examiner with a coefficient below 0.7 was replaced. Wiseman also produced essay reliability (stability) coefficients by correlating the aggregate marks from the first marking with the remark aggregates, and these values were 0.946 in 1943 and 0.91 in 1948.

The important point to note is that researchers were quite aware that individual markers vary within themselves. Underlying their studies at this time was one important assumption: there is no way of changing the nature of each grader's perceptions and judgements. Instead, the context in which the marker worked was controlled through manipulation of variables external to any one marker - the team he was placed in, for example, the pooling

of scores, and the specific instructions given to him by the investigator as to what method of grading and what scale to use.

In short, graders were studied in relation to other graders, to particular systems, or to stability over time. A grader was seen in a context that extended beyond himself. Even when studies required him to use different categories and criteria for making judgements, the focus of the researcher was on how reliability coefficients had been affected, not on the nature of the perceptions themselves.

Yet, there was opportunity at that time to examine those characteristics of graders that made them inconsistent, and to try to identify those graders whose judgements varied considerably within the group and within themselves. But the fact is that the use of mean scores and the phenomenon of statistical regression hid errant judges from view. Multiple-marking, as in the Devon System, allowed a grader to be consistently harsh, or consistently lenient, in relation to other markers in his team. He could even be inconsistent and undetected in a single marking operation, though it is unlikely that such inconsistency would have remained unnoticed in mark-remark situations that Wiseman rightly used. But the fact remains that, given a single marking session in which a grader pooled his marks with others, he could remain substantially unpredictable because the mean or the aggregate of the team formed the estimate of the true worth of each essay.

Of course, this was the reason why the multiple-marking was used. The notions of "true" score and errors of measurement from classical test theory acknowledged that error variance would operate within each grader, but that the mean of their judgements would be the best estimate of essay quality, and that error variance would balance across markers. But what if multiple-marking were not used? The fate of the candidate would depend on who marked the paper. And what if a complete team were invalid in its marking? The fate of the candidate could depend on which team marked the paper.

Gosling (1966) brought much of the research knowledge to bear on his large-scale study. Using papers and graders in the Commonwealth Secondary Scholarship Examinations in Australia over two years, he provided graders with model compositions for five points along a 15-point scale. Like Wiseman earlier in Devon, he was the Chief Examiner. Again, like others before him, Gosling felt that attempting to standardise the variance of each marker on each of the three writing tasks would be too time-consuming and laborious. Instead, he tried to solve the problem "by the nature of the marking system itself" (1966, p 39). After briefing and training the graders, and after controlling for halo effect (which he also put to the test at one stage of his investigations), he found correlations from 0.82 to 0.9 in 1964 and, in 1965, from 0.87 to 0.94, between the marking "system" and the Chief Examiner's marks. This procedure he used and put to the test as another indication of how "true" score might be considered by virtue of being an external criterion of the

quality of a composition. Correlations between system marks and examiner-mean ("true") scores ranged from 0.87 to 0.91 for separate writing tasks. When total weighted sums over all tasks were used, the correlation between system marks and examiner-means was 0.95.

From analysis of variance, he found no significant differences between sex of markers, nor between different times of the day when markers broke between marking sessions.

Despite close analysis of how correlations between individual markers and the mean-of-all-markers became over-estimated and under-estimated, according to whether the marks of the individual marker are or are not included in the calculation of the mean-of-all-markers, and despite the fact that he pointed out the limiting value of both procedures when using the pooling square (Thomson, 1939), he did not make it explicit which way he finally computed his coefficients.

He suggested that

> "... the best estimate of the correlation of an individual marker with the true score, when defined as a mean, is the mean of his correlations with all other markers divided by the square root of the mean of all inter-examiner correlations."
>
> (1966, p 15)

But the important point of all this is that his study was directed at manipulation of factors beyond the behaviour of any single marker; it was "outward-directed", focussing concern and attention on graders in such a holistic context that the individual inconsistencies of each grader were, in fact, obfuscated by the statistical procedures. He was concerned with finding "the reliability of the marking system as a whole (as distinct from individual marker reliabilities)" (p 41), no matter what variations might be going on within graders. In fact, the assumption underlying his study, typical of most studies in essay-grading, was that graders do vary within themselves, but, apart from controlling for halo effect and timing of sessions in a day, the internal validity of graders' perceptions could not be observed, let alone controlled. In fairness to his exhaustive study, it must be noted that he accounted for halo effect, but the control was exercised through manipulation of the papers, not through any substantial knowledge of grader-characteristics. Such was the state of knowledge and technique at the time.

Since the 1970's, the Chief Examiners in Australia have largely followed these earlier research findings in their attempts to establish high grader-reliability in English. So have informed teachers in their schools. Bennett and Hay (1971) described the public examination essay-marking procedures in all

Australian states. Rapid-general-impression marking is used; some multiple-marking; but mostly team-marking with pooling and consensus is used; halo effect is controlled; marking sessions are intense and as massed as possible to enhance consistency of grading over time; variance is interpreted as discrepancies amongst marker means.

In Western Australia, markers for the Tertiary Admissions Examination in English have been monitored, and individuals have been withdrawn from the panel if they are found to be too inconsistent. Bennett and Hay, as Chief Examiners in 1970, balanced "hard" and "soft" markers in the composition of teams of three (Bennett and Hay, 1971).

It must be emphasised at this point that the designs and statistical treatments of research studies up to the 1970's could not provide answers to the many questions about essay-grading which deal with the process of reading by the essay-grader.

Analysis of variance procedures and correlational studies could not account for the reasons for intra-grader inconsistencies. Per se, such statistical analyses do not permit relevant information to be revealed unless the variables themselves have been changed and alternative designs considered. The closest scrutiny of grader-characteristics had come only from factor analysis. Gosling (1966) reviewed the American study by

Diederich[10] as one of the most significant investigations into factors affecting the agreement of markers.

Diederich selected for the Educational Testing Service fifty-three markers from six different fields - college English teachers, social and natural scientists, writers, editors, business executives and law professors. After they had each marked three hundred essays from freshmen students sampled from three colleges, their grades were intercorrelated together with Board scores. The median correlation between all pairs of markers was 0.31; between all pairs of English teachers, the median correlation was 0.41. Apart from the writers and editors, the graders from any other professional field showed no more agreement amongst themselves than graders from different fields.

Factor analysis revealed five types of readers who each reacted mainly to one of the following qualities in writing: ideas, referring to clear, coherent, relevant and supportive material; form, meaning organisation and spelling; flavour, which referred to the novel, unconventional and interesting ways in which the personality of the writer came across; mechanics, such as grammar, spelling, punctuation; wording, involving diction and placement of words as elements of style.

---

[10] *Diederich, P. G. Factors in the judgement of writing ability. ETS Research Bulletin, 61-65. Princeton, N.J.: ETS, 1961*

These reader-factors were very similar in reference to the "categories" that Hartog (1941) had used as criteria on which his markers should grade. A consistent picture of how certain factors affect graders was now emerging. In general-impression marking, some graders judge more by certain qualities than other graders do. This is what Diederich found. So, presumably, the graders were looking for these characteristics in the essays themselves, and possibly doing so subconsciously. But Hartog's markers were consciously applying these criteria in analytic marking, one characteristic at a time. In both experiments, inter-marker inconsistencies were revealed, suggesting that, whatever grading system is used, the largest differences in grades arise because graders react differentially to the same writing qualities.

Here, then, were indications that differences in grader-characteristics can be identified in terms of the various writing-qualities that affect graders' judgements. All the above studies were examining differences between graders. Clearly, graders differ in their reactions to certain writing-qualities. The need was now apparent to examine how these variables might affect graders differently from essay to essay. It could well be that such "within grader" effects could account for some of the differences operating between graders.

At last, in recent years, there has been some research that has manipulated writing-qualities as independent variables in order to observe the effects on graders. Hake and Williams (1975,

1979) have reconceptualised the variables, placing the emphasis on the specific judgements that graders make in reaction to essays that have certain characteristics built into them. Together with this reconceptualisation of variables, they have used some new statistical procedures that allow precise measures to be made of both graders and essays.

Because of the immediate relevance to the present study of Hake and Williams' research and because this study uses some of the new statistical procedures, Chapter III will be devoted to reviewing two studies by Hake and Williams which provide the basis for the design of the present study.

CHAPTER III

THE STUDIES OF HAKE AND WILLIAMS IN ESSAY-GRADING

Hake and Williams conducted a pilot study in 1975 to investigate the effects of style on essay-graders. They took four papers which had been written for a writing-competency examination at Chicago State University and re-wrote them to provide four pairs of papers. Each member of each pair differed from the other only in its style. One version of each pair was written in what they called a "nominal" style, and the other version was in a "verbal" style. The nominal style was made by employing an abstract noun in place of a verb; the verbal style used a verb wherever possible in place of an abstract noun.

The eight experimental papers were included with twelve other unchanged essays and given to ten composition instructors from a variety of educational institutions to grade on a five-point scale. The graders were asked also to comment in writing about each essay that they marked.

Since there were ten graders and four pairs of experimental essays, there were forty opportunities for comparisons of pair scores. On thirty-five occasions, the nominal style papers were scored higher than their verbal style counterparts. On the remaining

five occasions, the two versions of a pair received the same score. The authors emphasised that the verbal style is a more direct, concrete and vigorous form of expression, and more economical in words. Yet the verbal style was never rewarded with higher grades. In addition, the best of the four original papers, when it was in its verbal version, rated lower than the worst paper in its nominal form when the scores for the two papers were summed.

From these results, Hake and Williams suggested that what composition teachers say they value in good writing and what they do value when they judge essays are not the same. From the written comments, it appeared that graders perceived contradictory qualities, even when they were marking the same essay in its different versions; for some essays, graders even perceived errors that did not exist.

Though some of their implications and suggested causes for these surprising results can be questioned, Hake and Williams had revealed an area for more research by virtue of the new approach in design. Both the findings and the design provided a foundation for the development of the present study.

In 1979, Hake and Williams conducted four experiments along similar lines.

Experiment 1:

Five high school and five senior college teachers graded twenty papers within which were four pairs of essays. One essay of each

pair was written in nominal style, with verbs and adjectives transformed into abstract gerunds and nouns. The writing was characteristic of the inflated, indirect prose of officialese. The other essay of each pair was written in verbal style, with subject-verb-object grammatical strings to correspond with an action-goal structure.

In all other respects, the two essays of each pair were identical. Errors were deliberately created in spelling, punctuation, grammar and usage, and were placed in identical positions. Paragraphing, content, organisation, handwriting and even marginal scribblings were identical.

Using a five-point scale, ten graders rated by general-impression. They also made written comments about each essay that they read.

The preference for the nominal versions was overwhelming. Only in fifteen percent of the ratings did markers award the same grade to both versions; all other ratings showed that the nominal version was preferred over the verbal version. The syntactic maturity of the nominal papers "seemed sufficient to trigger judgements of a paper's profundity and of the superiority of its logic and organisation" (1979, p 7). In addition, the difference between nominal scores and verbal scores for the essays that were of higher quality was larger than the difference between nominal and verbal scores for the poor quality essays.

Hake and Williams make the following statement about the results of the first experiment:

> "We must entertain the possibility that many teachers respond in the same way toward their own students, encouraging precisely the stylistic values we claim we reject and discouraging precisely the stylistic values we claim we support. And as a consequence, our students acquire - or at least are not pragmatically encouraged to avoid - the stylistic values that we claim we deplore in adult prose."
>
> (1979, p 9)

Experiment 2:

More papers and more graders were used in a replication of Experiment 1. The authors suspected a Hawthorne effect, and noted that forty papers in less than four hours was too much a marking burden to ensure consistent judgements. In the first experiment, graders had been able to work on the papers at home in their leisure, ten one fortnight and ten another fortnight. The second experiment had the thirty-six teachers mark at the University of Chicago on two Saturday mornings two weeks apart.

While more of the nominal versions received the higher grades in each pair than the verbal versions, the differences were not significant statistically. The authors noted that the verbal papers were scored higher only 30% of the time, and this could be interpreted as significant if one expects that concise, direct writing should be judged with more approval 100% of the time.

Statistical significance was found, however, when marks of high school teachers were compared with those of junior and senior college teachers. The high school teachers preferred the nominal versions significantly more often than college teachers did. Also, good nominal papers were scored higher than good verbal papers, but amongst the poor quality papers, verbal versions scored significantly higher than the nominal. This type of result was significant for high school teachers and the senior college teachers, but not for the junior college teachers:

> "The least sophisticated teachers of composition, high-school teachers, were the most impressed with what appeared to be verbal sophistication. Regardless of the intrinsic quality of an essay, a nominal style so strongly bespoke verbal dexterity and implied intellectual maturity that the high school teachers responded to a nominal style as evidence of intrinsic quality. The senior college teachers were more discriminating, but that very discrimination reveals how powerfully style alone can influence judgements. Apparently, a nominal style and intrinsic quality re-enforced one another: profound style = profound content. But they perceived a nominal style in an intrinsically bad paper as a mismatch, as an attempt to cover up an absence of substantial content with pretentiously inflated language. The more balanced response by the junior college teachers may indicate more experience teaching composition, less of a willingness to be influenced at all by pretentious language."
>
> (1979, p 12)

These implications have been quoted at length because there are contradictory assumptions and inferences within them that will be taken up later in this chapter.

Experiment 3:

Essentially, the same experimental design was employed, this time in Georgia with papers from a Regent's examination in writing. Thirty-six graders under the supervision of persons other than Hake and Williams graded their papers as usual, but inserted into the batches were again two pairs of essays, one pair of a better quality intrinsically than the other pair, and both pairs consisting of nominal and verbal versions of otherwise identical responses to the examination question. Statistical significance was not reached again, even though the trend was strongly in favour of awarding higher grades to the nominal versions; 29.2% of scorings awarded nominal papers superior to verbal, while 12.5% of ratings awarded the higher grades to verbal versions. On average, 58.3% of ratings awarded the two versions the same grade. With some further analyses, Hake and Williams tested their hypotheses by stratifying the graders according to range of grades they used and according to whether they were Junior or Senior College teachers. No significant differences resulted, though trends were similar to the earlier experimental findings.

Experiment 4:

In this experiment, Hake and Williams employed new statistical procedures which will be explained in some detail because the present study uses similar ones.

At first Hake and Williams used twelve graders to grade essays on a four-point scale by general-impression. Comparisons were made of the mean values awarded to nominal and verbal versions of the essays, and these mean values were calculated for high school teachers, junior college teachers and senior college teachers. The resultant chi-square was significant at the 6% level only. Once again, there was a strong trend to favour the nominal papers, and the high school teachers did so significantly more than the other teachers.

The twelve graders were also trained in how to mark using Hake's Marking Frame (Hake, 1973). The frame directs graders to specific types of flaws that can occur in written communication, and the grader simply records each flaw as it appears for him. The recording of flaws from grader to grader does not have to be uniform, as different graders will perceive flaws or not according to their individual interpretation of the meaning of the communication and the student's competence in technique. Even where two graders do perceive a flaw, one might record it as a flaw in the dimension of logic and sentence-structure, while the other might record it as an error in mechanics, such as punctuation. Altogether, Hake developed four dimensions for categories of flaws. The tallying of these flaws indicates the strength or weakness of the essay as a piece of purposeful and effective communication.

The important characteristic of the use of Hake's Frame is that each observation of a flaw is a discrete event, independent of the observation of any other flaw. The grader does not have to retain any cognitive scheme of criteria other than identifying any hindrance

to the writer's communication, and so, in this sense, the marking is objective because discrete observations are made independently one at a time on a single criterion - flaws in communication for the reader.

The graders, then, graded the essays, and their responses in terms of flaw-counts depended on the nature of the essay they were marking at any one time. In other words, the frame was not the independent variable in the way that, in earlier research, workers had made the marking system the independent variable.

By using a development of the Rasch Multiplicative Poisson Model (Andrich and Hake, 1973), the flaw-counts of graders were transformed into measures of grader-harshness on a linear interval scale, that is, the graders were calibrated. Then, adjustments were made to each grader's flaw-counts according to his harshness of grading so that these "standardised" counts were directly comparable. Now the results of the effects of stylistic changes in essays stood out:

(a) The high school teachers found more errors on each dimension of the verbal papers, and they had graded nominal papers higher.

(b) The junior college graders found more usage and mechanics flaws in verbal papers, even though, in total, about the same number of flaws was recorded

for each of the two versions of essays; but they still favoured nominal versions over verbal versions in their overall grades.

(c) The senior college graders also favoured the nominal versions on overall grades, but their flaw counts were higher for nominal style than for verbal style. Like the junior college graders, their flaws for nominal papers were on the logic dimension, while the verbal versions attracted their attention more towards the mechanics dimension.

It seems clear that nominal style suggests intellectual maturity, and, as Hake and Williams point out, graders react by perceiving flaws in logical organisation and coherence of meaning. Verbal style seems to suggest less intellectual maturity, and so the grader perceives flaws at a mechanistic level. The authors suggest an explanation:

> "... it may be that a nominal style is in fact cognitively more difficult for the grader to follow ... And so the reader responded to that lack of clarity by criticizing the paper's cognitive and stylistic dimensions. <u>But the nature of that difficulty, a seemingly mature prose style, was still sufficient to trigger the overriding judgement that evaluated the nominal paper as relatively superior</u>. If this is true, then the verbal papers were too clear for their own good."
>
> (1979, p 20)

Experiment 5:

To test whether the nominal style is more difficult to process cognitively, even at a simple mechanical level, Hake and Williams employed typists to transcribe the essays. Typists were found to type faster and with fewer errors when their copy was from verbal versions rather than nominal versions of printed material. The more experienced and professional typists were less influenced than high school typists in these ways, though no statistical tests were reported from the study. The point the authors made here was that nominal style is, in fact, more difficult to process, and they drew from the literature of psycholinguistics to reinforce the conclusion.

Overall, Hake and Williams have indicated the following characteristics of the encounter between grader and essay as a consequence of their reconceptualisation of variables and their use of new measurement models.

(a) Some graders are consistently harsh or lenient - and these characteristics can be measured.

(b) Essays with certain stylistic components cause graders to award grades inconsistent with both the quality of the essay's other components and with their own grading trends.

(c) Substantial numbers of graders, numbers approaching statistical significance, are affected in the same ways with regard to the grades they award when essays contain different stylistic devices.

(d) When certain stylistic devices are changed in an essay, the grader tends to perceive different competencies in the essay, competencies which are not related to style. In fact, graders tend to imagine higher and/or lower quality in dimensions of essays which should, one would expect, not cause such a shift in the grader-reaction because they do not concern style as such.

(e) Essays of high quality tend to receive higher grades when written in heavy, formal, indirect style, than when they are written in simple, direct style. The grader tends, also, to perceive flaws more in logic and coherence when the indirect style is before him.

(f) Essays of low quality tend to receive higher grades when written in simple, direct style. The grader tends, also, to perceive flaws more in usage and mechanics when the direct style is before him.

(g) An essay is more likely to fail when it is written in simple, direct style, and the grader is likely to imagine that there are more flaws in it than there actually are.

(h) Most graders, irrespective of professional levels of teaching, seem to be affected in varying degrees.

(i) Methods of grading, that is, general impression and using analytic frames, do not prevent graders from being influenced in the same unexpected ways by stylistics.

Some of the implications and explanations offered by Hake and Williams (1975, 1979) for the above characteristics of essay and grader encounters need closer examination. It is possible that one would expect children and students to be taught to write clearly, concisely and directly; it is also possible that some people want children and students to be taught to write using official-sounding style. It is possible that some people want children and students to be taught to do both as the need arises. It may, therefore, be an over-generalisation to state, as Hake and Williams (1975) have, that teachers tacitly subscribe to the use of indirect, formal style while overtly they advocate clear, direct style. This implication goes beyond the strict findings and reasonable implications of their two studies because the overt values of teachers themselves were not investigated. It could well be that Hake and Williams' assumption that teachers do teach for direct style in writing is tenable, but with no evidence otherwise shown, it could be just as tenable an assumption that teachers are trying to develop children's control of syntactical, lexical and grammatical structures towards the

distant goal of "maturity" of expression. In fact, as has been shown in the review of English curriculum literature, this is happening in composition instruction, particularly in the United States of America.

CHAPTER IV

DESIGN OF THE STUDY

The hypotheses and design of the present study are derived essentially from the 1975 study by Hake and Williams who re-wrote students' essays so as to build in certain stylistic qualities. They then studied what effects these qualities had on graders' judgements. Like both studies of Hake and Williams (1975, 1979) described in Chapter III, this study examines whether graders reward essays written in simple, direct style more than they reward essays written in more complex and indirect style.

This study refines and extends the stylistic qualities that Hake and Williams used as independent variables. Specifically, the present design involves two independent variables: Voice and Diction. These variables were incorporated into students' essays by re-writing the essays with no changes other than the concentrated use of a particular voice and a certain choice of words. How these two variables were employed is described more explicitly below, but certain definitions of concepts and terms must first be stated.

The terms "graders" and "markers" are used synonymously throughout this dissertation to refer to persons who assess the value of an essay. Strictly speaking, graders award a grade, such as a letter or number, while markers actually award grades and correct flaws on the papers. Unless stated in contexts that might be otherwise ambiguous, the two terms are used to represent the single act of awarding a grade to represent the quality of an essay.

By "essay" is meant an extended piece of writing typically undertaken by secondary school students in response to some topic. The word "composition" may also be used, and perhaps more appropriately in regard to the writing of school children. The terms can be used synonymously in the context of this study, even though the essay, as a literary form, is typically longer, more discursive, and purports to be of significant interest for a universal audience.

Elements that make up style are such characteristics as the writer's choice of words, his use of figures of speech, his control of sentence-structure in various ways, his point of view (that is, the position from which he views his material), and the selection of detail (Strunk and White, 1972). In the present study, a "style" is the result of manipulating the two variables, Voice and Diction.

Strictly, style ought to be the result of a writer's controlled synthesis of all the above elements. Simply to change diction without modifying other elements is not to change style properly at all. Rather, it is a change in a stylistic device. For the purposes of experimentation, however, Voice and Diction had to be manipulated in this manner to produce the versions of essays required. From a literary point of view, to effect such changes in stylistics is crude and arbitrary, and it is artificial to generate essays so consistently of one voice and with restricted choice of words.

Accordingly, the use of the word "style" in this dissertation signifies the particular use of stylistic devices.

The stylistic devices were incorporated into essays by manipulating sentences into either the active voice or the passive voice, and then the words of these sentences were of either Anglo-Saxon origins or of Greek and Latinate origins. The four styles thus generated are represented in Table I.

TABLE I

STYLE CHARACTERISTICS OF RE-WRITTEN ESSAYS

	Saxon Diction	Latinate Diction
Active Voice	Style I	Style II
Passive Voice	Style III	Style IV

Anglo-Saxon words are usually shorter than Greek and Latinate words and are more vigorous and direct. Greek and Latinate words are usually longer, often more euphonious, and less likely to be used in everyday language. They are more official-sounding and less vigorous in many contexts in which they are used (Orwell, 1946; Boulton, 1954). In this study "Saxon" and "Latinate" are used as short-hand terms to indicate these changes in diction.

The Active voice is evident when a subject-verb-object grammatical string is used for the sentence-structure. This direct, transitive use of verb provides a clear action-goal structure that makes for directness, simplicity and vigour of expression. The Active voice requires actions to be described using verbs, whereas the Passive voice describes actions using nominalisations of verbs, particularly in linkage with clauses in the third person conjugations such as "there was ... " and "it has been ...". Where verbs are used, they are intransitive and the effect is less vigorous and less direct. Two examples will indicate the use of Active and Passive voice in this study:

ACTIVE: Everybody shouted when the team scored the goal.

PASSIVE: There was shouting from everybody when the goal was scored by the team.

ACTIVE: He keeps up payments or loses the goods.

PASSIVE: The payments must be kept up or the goods are lost.

The combination of one voice, such as the passive, with one type of diction, such as Latinate, produces a style that is very official-sounding, ponderous, academic in sound and sentence structure, and, in some writing, elegant because it is not vigorous. It is a style very different from one where the active voice is combined with Saxon diction. The question is: What effects will various combinations have on essay-graders?

Hypotheses

In general, it is hypothesised that graders will award significantly higher grades to any essay written in Latinate diction than they will to the same essay written in Saxon diction. Similarly, when an essay is written in Passive voice, it is hypothesised that it will receive significantly higher grades than when it is written in Active voice.

There are four possible combinations of Voice and Diction that have been incorporated into the re-written versions of students' original essays. There could be significant differences in grades amongst these four combinations. There could also be differences between the original essays and any of the stylistic combinations. It is difficult to estimate which style will attract grades most similar to the grades awarded the original essay; most students write in an individual manner which draws upon the use of voice and diction according

to the purpose for writing, their grasp of the subject, and their own personalities. In fact, most people write in this way.

Formally stated, the null hypotheses for this study are as follows:

1. There is no difference between the grades awarded to essays written in Active/Saxon style and the grades awarded to the same essays written in Passive/Latinate style.

2. There is no difference between the grades awarded to essays written in Active voice and the grades awarded the same essays written in Passive voice.

3. There is no difference between the grades awarded to essays in Saxon diction and the grades awarded the same essays written in Latinate diction.

4. There is no difference between the grades awarded to students' original essays and the grades awarded to the same essays when re-written in any of four styles, namely Active/Saxon, Active/Latinate, Passive/Saxon and Passive/Latinate.

These hypotheses were put to the test experimentally in various phases of the study. The first phase was a pilot study in 1976, and the other two phases were conducted in 1977 and 1978.

1976 Pilot Study:

Four experienced and, at the time, current examiners for the Western Australian Tertiary Examination in English were selected in the first part of the pilot study, and another four for a replication of the pilot study. None of the eight examiners had had less than three years' experience as a marker on the panel. All but three were senior masters of English in senior high schools, including one from a private Catholic school. Of the three who were not senior masters, one was a teachers' college lecturer in English with primary school teaching experience, one was a specialist English teacher in a high school, and the third was an English Curriculum Advisor in the Curriculum Branch of the W.A. Education Department, having been seconded from secondary schools.

All were paid to grade and comment in writing upon twenty-five essays in a single marking session with no set time-limit. The session was held in a seminar room in one of the high schools.

1977 Phase:

Graders were sixty third-year teacher-training students (final year students) who had voluntarily enrolled in the

investigator's elective course in classroom measurement and evaluation at Mount Lawley College of Advanced Education. About a half of them enrolled in the first semester, and had had up to eight weeks of practice-teaching in schools to that stage. The other half enrolled in the second semester, studying the same measurement and evaluation course which was a repeat of the first semester unit. These latter students had had twenty weeks of practice-teaching experience, including a recent programme over twelve weeks as assistant teachers in the schools.

The sixty subjects were obviously not experienced teachers. They had felt the effects of an English curriculum only three years earlier as secondary school students completing their twelfth year of schooling. They were probably very much aware of current practices in written-expression instruction from the practical point of view as receivers of such instruction. In preparing for the Tertiary Admissions Examination in English less than three years earlier, they would have been subject to teachers' overt statements of criteria and judgements in the evaluation of their writing ability.

More recently, they had been instructed in College courses in English and how to teach it in primary schools. Virtually on the eve of their taking up professional appointments as teachers, they would have been well-acquainted with the latest trends in English curriculum and methodology, and aware of criteria involved in the evaluation of compositions.

The investigator was careful not to brief these students in any aspect of essay-grading research or practice. The knowledge they had accrued of the subject would have come only from other courses and other departments in the College and from experience in schools. It could be reasonably assumed, then, that they represented the new teachers in the field, fairly informed about new curriculum trends and close to the implementation of primary and secondary English curricula.

The sixty subjects were required to grade and comment in writing upon twenty-five essays in a single marking session of two hours. For this participation, they were granted a full ten percent of the assessment requirements for their measurement and evaluation unit. They were keen to receive this award because they were not required to prepare formal assignments to earn the ten percent. From the investigator's observations, they were motivated to grade, and, because they were told that they were taking part in an experiment, they were clearly interested and involved.

1978 Phase :

Fifty-seven student-teachers were used as graders. In all respects, they were the same sort of subjects as those of the 1977 group.

Procedures

1. Five sample essays from a Year 12 English class at a suburban senior high school were selected to represent a range of writing ability. The essays had been written in a single lesson in response to one topic that was announced immediately before the students were left to write: "'Australians are becoming soft.' Discuss." No class discussion preceded writing.

2. The five essays were re-written by the investigator in each of the four styles described earlier (see Table I, page 46), thus generating twenty experimental essays. In addition, the essays were typed, and original errors in the students' essays were transcribed. Spelling errors often had to be translated to new words, and it was found necessary on some occasions to create a spelling error of the same type as the original error and to incorporate it into the new word, or place it as close as possible to the contextual location of the original. Samples of these original and experimental essays are included in Appendix A, together with notes on techniques used to overcome etymological problems.

3. The twenty essays were presented to the graders to grade, together with five "padding" essays which had been simply typed without changes from the originals. These five had been chosen at random. The twenty-five essays were randomised within patterns such that each of the Styles, I, II, III, and IV, was "rotated" in the order in which it appeared amongst graders. The original student-writer represented in the experimental essays was also "rotated" so that each one of the authors had one of his versions read first by some marker. With the randomisation also of the occurrence of the padding essays, every effort was made to control extraneous variables that might have operated as a result of essay order on the graders.

4. Markers graded each essay on a five-point scale, A, B, C, C- and D, the two low points being identified as Fail/ Unsatisfactory for Year 12 students in Term I of their final year. The graders were required to award grades on three dimensions: the first was to be an Overall grade resulting from general impression of the paper; the second was to be for Mechanics and Usage; the third grade was to be an evaluation of the Logic and Organisation of the essay. They were encouraged to make their own comments in writing about each paper. Specific instructions given to the graders are shown in Appendix B.

5. In the case of the 1978 phase of the study, the graders were required to grade the original student essay as well as its four versions. This would have meant grading thirty essays in the single marking session of two hours. As a safeguard against the possible effects of fatigue from too large a marking load, one of the original essaysand its four generations was removed from each grader's batch of papers. The particular set of essays thus removed was "rotated" amongst the graders, so that, in fact, all five sets of experimental and original essays were graded, but not all graders graded the same sets - one set was removed for sub-groups of graders. All essays were graded, in effect, by about forty graders each.

All graders graded all padding essays as before.

Statistical Treatment and Analysis

The purpose of this study required that grades given to the various versions of the essays be compared to see if graders' judgements had been affected by the use of style. But before the effects of the four styles were analysed, a close check was made on the reliability of the 1978 graders. This was considered necessary to ensure that differences in grades were due to the treatments (styles) and not to inconsistencies in the graders as a result of extraneous influences. The reliability of the 1977

graders was also noted during the particular procedures employed as an integral part of the statistical treatment.

After the check for reliability, the data from each of the three phases of the study were analysed individually for effects of style. The statistical design used was multivariate repeated-measures analysis of variance. This was the natural choice of design, since each grader had awarded a total of sixty grades; each grader graded twenty experimental essays and each essay received a grade on the dimensions of Overall Quality, Mechanics and Logic. The analysis of variance of repeated measures permitted differences between grades to show as the graders moved from style to style. Furthermore, the differences amongst the grading dimensions between styles could also be analysed by using multivariate design.

The letter grades awarded by the graders had to be converted to numerical equivalents, 0 to 4, but these grade-scores were not used as such in the statistical analyses. The grade-scores were transformed to a linear interval-scale by the use of the Rasch Polychotomous Rating Model as adapted by Andrich (1975, 1978a, 1978b). Details of this model are included for reference in Appendix C, but a brief description now follows.

Grade-scores on a five-point scale do not provide much in the way of variance, yet it is the variance in scores which indicates differences between treatments. A finer measure of grader judgements would allow subtle variations and trends to show.

By using the Rasch Rating Model, more precise measures of writing ability of each essay were possible because the five-point scale-scores were translated logarithmically onto a linear scale which allowed more appropriate spacings based on the ratings. This transformation made the use of analysis of variance more justifiable than it would have been with the use of ratings directly.

The Rasch model actually calibrates the essays in terms of "ability" along a continuum and it provides known degrees of precision in doing so. The model also calibrates the graders in terms of harshness of grading along a continuum.

With more precise measures of the worth ("ability") of each essay as it appeared in its various stylistic versions, the analysis of variance was carried out. And with more precise measures of harshness of each grader, it was possible to observe how closely each grader graded in relation to himself and to the other graders. Such a reliability check was far more stringent and precise than that provided for by traditional procedures.

With the 1976 pilot study, graders' written comments were interpreted in accordance with statistical outcomes as far as possible. With the 1978 data, essay ability measures of the original unchanged essays were also incorporated into the analyses.

Choice of Stylistic Devices:

Only voice and diction were used as independent variables. There are other devices and elements of style that may be put to the test. For example, the use of figures of speech such as metaphor and simile could be tested; the effect of different points of view could well provide significant effects on graders' judgements as essays are written in third person or first person. Other variables are also possible, such as sentence-structure, organisation, and use of concrete or abstract examples.

Choice of Writing-Task:

All essays were on the one topic which required a response that limited students to writing a discussion, exposition or argument. While this is typical of the sort of writing-task required of upper-secondary students, it limits the generalisability of results, since wider writing contexts that involve, for example, expressive and poetic writing with complete freedom of choice of topic might affect graders of such compositions differently.

However, by limiting the choice to a single topic, the study controlled extraneous variables that might operate on graders had there been more variety in content and subject.

It has been pointed out in Chapter II that graders are more likely to be reliable amongst themselves when the writing task is clearly and uniformly applied to all students.

Selection of Samples of Graders:

One interesting outcome of the study will be the extent to which inexperienced student-teachers grade in the same way as experienced teachers and public examiners. There were only eight public examiners who were used in the study, though this represented about twenty-two percent of the entire panel of examiners for T.A.E. English. Some caution might be exercised in attempting to generalise findings to all English examiners, but such caution would be due to the small size of the sample rather than to bias in its composition.

Nature of the Experimental Essays:

In re-writing student essays so as to build in the various stylistic devices, the investigator produced essays that were each inflexibly and completely in one "style" as far as voice and diction were concerned. To an extent, this is an artificial control of the independent variables, since it would be expected that no writer writes so exclusively in one voice and with conscious bias in vocabulary. Yet the results of the Hake and Williams' research would indicate that this is not a serious limitation. Hake and Williams noted that, in fact, most graders

did not know that the same original essay was the basis for the generation of experimental essays. This absence of awareness by the graders would have been due to the nature of the experimental design. Nevertheless, it is pertinent to point out this characteristic of the experimental essays as a limitation insofar that it could produce effects in the grading behaviour that are subtly extraneous and uncontrolled. Such effects could well be frustration, or irritation in the minds of different graders who might well award a particular grade partly as a consequence of these affective states rather than as a consequence of the independent variables alone.

The above limitations in the use of experimental essays and the samples of graders reveal that there are some assumptions that have to be made clear at this point.

Assumptions

1. That the graders in the pilot study were motivated to read and grade the essays to the best of their ability. This assumption can be defended on the grounds that the eight graders were paid for their single marking session.

2. That the teacher-trainee graders were likewise motivated. This assumption can be defended only on the grounds that the operation was conducted as a class exercise for which ten percent of the assessment for their measurement and evaluation courses was granted them for applying themselves to the

task of grading. These subjects were also aware that they were taking part in an experiment, and this is likely to have motivated them to apply themselves conscientiously to the task.

3. That the original essays represented the full range of writing ability in final-year secondary students. From the investigator's experience as an English teacher, and from the grading data, this assumption holds.

CHAPTER V

RESULTS AND DISCUSSION

The main results for this study arise from the analysis of variance procedures by which the ability scores for essays in different styles were compared. These ability scores were derived from the graders' original grade-scores and transformed to a linear scale by employing the Rasch Polychotomous Rating Model as described in Chapter IV and Appendix C.

The Rasch model is a probabilistic model, and part of the procedures entails calculating the initial estimates of the subjects' ability measures, refining them according to the spread of scores and persons, and then predicting back from these corrected estimates to see if the observed data fit the model. If there are wide discrepancies between expected values from the model and the actual observed data, then there is clear evidence that such observed values are inconsistent beyond simple error variations. 95% confidence limits were chosen throughout this study.

This checking to see if the data fit the model is standard procedure. In the process, the reliability of subjects' responses is examined. If any subject's observed responses do not fall within the confidence limits anticipated from the model, then that subject is removed from the analyses.

Reliability of essay-graders in this study was checked in this way for the large-scale phases in 1977 and 1978, but not for the initial pilot study in 1976. Only the 1978 analysis of fit is reported here. The 1977 analysis of fit is included for reference in Appendix C. Both analyses found all graders to conform to the model.

It was necessary to consider the reliability of the graders carefully because it had to be evident that they were reacting to the stylistic versions of the essays as "treatments"; they could not be inconsistent as a result of the influence of extraneous factors. Once their fit to the model had been confirmed, then analysis of variance for treatment effects could be conducted.

Reliability of 1978 Graders

Of 57 graders, 52 provided full and complete responses. Since these graders had each graded all of the padding essays, but not all the experimental essays, their grade-scores for the padding essays were used as the basis for establishing their harshness measures.

The application of the Rasch Polychotomous Rating Model followed the procedures established by Andrich (1978a, 1978b). Initial estimates of grader harshness were calculated for each grader on each of the three grading dimensions separately. Two class intervals of those estimates were formed so that graders were grouped into [illegible] or lenient in their grading on each dimension.

Predicting from the model back to the observed data, it was possible to aggregate the discrepancies (residuals) between observed and expected proportions of graders awarding each of the grades on the five-point scale. These discrepancies should show no systematic pattern for any essay. A chi-square test on the observed and expected proportions responding to the five grades for each essay was carried out, and the results are shown in Table II.

TABLE II

CHI-SQUARE VALUES FOR OBSERVED AND EXPECTED PROPORTIONS OF GRADER AWARDS TO EACH GRADE-SCORE ON 1978 "PADDING" ESSAYS

Padding Essay Code No.	Ability Logit Estimate	Lenient Graders' Mean Logit	Harsh Graders' Mean Logit	Chi-square (df = 0.80)	Significance (p<)
P21 O*	-0.364	-0.215	0.592	0.363	.459
P21 M*	-0.341	-0.331	0.836	1.829	.135
P21 L*	-0.348	-0.229	0.711	0.281	.506
P22 O	-0.236	-0.215	0.592	0.085	.686
P22 M	0.019	-0.331	0.836	0.107	.656
P22 L	-0.398	-0.229	0.711	0.331	.476
P23 O	0.579	-0.215	0.592	0.169	.590
P23 M	0.381	-0.331	0.836	2.585	.081
P23 L	0.461	-0.229	0.711	0.169	.591
P24 O	-0.473	-0.215	0.592	0.576	.367
P24 M	-0.501	-0.331	0.836	0.670	.336
P24 L	-0.156	-0.229	0.711	0.998	.252
P25 O	0.594	-0.215	0.592	1.127	.227
P25 M	0.441	-0.331	0.836	0.440	.421
P25 L	0.441	-0.229	0.711	0.240	.624

*O = Overall grading dimension
M = Mechanics grading dimension
L = Logic grading dimension

Joint tests of fit for graders' responses on each grading dimension produced chi-square values of: 2.320 for Overall (df 4, $p<.677$); 5.631 for Mechanics (df 4, $p<.228$); 1.913 for Logic (df 4, $p<.752$). This shows that the graders were consistent as a group; their responses to essays conformed to expected probabilities derived from the model. There were no erratic inconsistencies amongst them, though this does not mean that they did not sustain different levels of harshness. If there were differences, then these were consistent.

In all phases of the study, the same consistency of grading was observed. No graders had to be withdrawn from the subsequent analyses because of any significant inconsistencies amongst graders or within an individual grader.

The effects of the stylistic treatments were then analysed by analysis of variance, following the refinement of essay and grader parameter measures (Andrich, 1978b).

1976 Results

The transformed grade-scores for each essay were averaged and collated in the repeated measures format as shown below in Table III. The means for each essay are shown as rounded to the nearest whole number. The values for each essay on the linear scale that result from the Rasch Model procedures have been centred around 50 for

convenience rather than around zero. This was done by multiplying each value by 4.55 and adding 50, a procedure that does not alter the relative values of essay awards since the linear scale onto which they are transformed by the model is an equal-interval scale.

TABLE III

MEANS OF TRANSFORMED SCORES FOR EACH ESSAY ON EACH GRADING DIMENSION : 1976 DATA

	Style I Active/Saxon			Style II Active/Latinate			Style III Passive/Saxon			Style IV Passive/Latinate		
Essay	O	M	L	O	M	L	O	M	L	O	M	L
V	57	58	57	63	66	58	52	51	52	63	59	58
W	52	50	53	58	54	58	56	53	57	65	59	62
X	49	47	49	53	54	52	47	44	48	53	52	52
Y	41	37	46	45	42	48	40	40	46	43	40	48
Z	42	43	46	49	44	49	42	39	44	50	48	51
Means: O	48.2			53.6			47.4			54.8		
M		47			52			45.4			51.6	
L			50.2			53			49.4			54.2

For ease of inspection, the means for each style are presented as "cell means" in Table IV.

TABLE IV

CELL MEANS OF TRANSFORMED SCORES FOR ESSAYS BY STYLE ON EACH GRADING DIMENSION (1976 DATA)

		DICTION				Summed Row Means	
		Saxon		Latinate			
VOICE	Active	O	48.2	O	53.6	O	101.8
		M	47	M	52	M	99.0
		L	50.2	L	53	L	103.2
	Passive	O	47.4	O	54.8	O	102.2
		M	45.4	M	51.6	M	97.0
		L	49.4	L	54.2	L	103.6
Summed Column Means		O	95.6	O	108.4		
		M	92.4	M	103.6		
		L	99.6	L	107.2		

The first hypothesis to be tested was the interaction between the Diction and Voice factors. This tested whether the difference between the means of the Saxon and Latinate essays in the Active voice was the same as the difference between the means of the Saxon and Latinate essays in the Passive voice. The interaction effects over the whole table were found to be not

significant; the analysis of variance summary table is shown in Table V.

TABLE V

ANALYSIS OF VARIANCE TABLE FOR INTERACTION EFFECTS OF VOICE AND DICTION (1976)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
Overall	20.0	4.0	.1162	(1,4)
Mechanics	7.2	.3318	.5955	(1,4)
Logic	20.0	3.6364	.1293	(1,4)
		MULTIVARIATE F		
		.8543	.5791	(3,2)

The insignificance of the interaction effect means that any differences in diction were the same, irrespective of voice; similarly, any differences in voice were the same, irrespective of diction. For graders in general, this could mean that if a significant difference were to be found in Diction, then grades could vary significantly from Saxon to Latinate, no matter whether the essays were in active or passive voice. Likewise, if a significant difference were to be found in Voice, then graders could be awarding significantly different grades to Active as against Passive, no matter what diction was being used.

The second hypothesis tested one of the main effects, that is, the Active against the Passive voice. With an insignificant interaction, it was meaningful to test this hypothesis. It was anticipated that the Passive Voice would receive significantly higher awards than the Active.

The summary table follows in Table VI .

TABLE VI

ANALYSIS OF VARIANCE TABLE FOR MAIN EFFECTS OF VOICE (1976)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
Overall	.8	.0201	.8942	(1,4)
Mechanics	20.0	.3008	.6126	(1,4)
Logic	.8	.0359	.8590	(1,4)
		MULTIVARIATE F		
		1.1927	.4863	(3,2)

Graders were not influenced significantly one way or the other in awarding grades to essays written either in active or passive voice. As shown in Table IV earlier, the summed means for Active over both types of diction were 101.8 (for Overall), 99.0 (for Mechanics) and 103.2 (for Logic), while the summed means for Passive were 102.2, 97.0, and 103.6 respectively. Evidently, the eight public examiners were not any more responsive to essays written with direct and simple syntax than to essays written with less direct, inverted and more cumbersome syntax. Apparently, they were equally influenced by either of the two forms of expression. This is surprising if the essays are examined with this finding in mind.

Samples from essays are shown below:

Essay W:

Active: "The housewife finds housework easy with the latest technical aids."

Passive: "The housewife has modern technical aids by which housework is made easy."

Essay X:

Active: "Moreover, in numerous instances by the time he has completely honoured his debt for one commodity he needs to purchase another and the vicious circle continues."

Passive: "Moreover in numerous instances by the time the debt for one commodity has been honoured in its entirety, the item must be substituted by another and thus the vicious circle is continued."

Essay Z:

Active: "Not many Australian men have the time or effort to keep up the way they should be ..."

Passive: "No time or effort is there for their image to be kept up by many Australian men ..."

It seems strange that the passive voice should be seen as not significantly different from the active. Under the conditions by which graders make their judgements, maybe the differences are not readily perceived. The differences seem obvious when one is afforded the time to read the samples analytically. Maybe time or some other factors of a social or psychological nature act against the graders and prevent opportunities for reacting to the differences that voice produces.

The third hypothesis tested was the effect of Saxon Diction against Latinate Diction. The latter was expected to receive higher awards. The summary table follows in Table VII.

TABLE VII

ANALYSIS OF VARIANCE TABLE FOR MAIN EFFECTS OF DICTION (1976)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
Overall	819.20	47.6279	.0024*	(1,4)
Mechanics	627.20	31.8376	.0049*	(1,4)
Logic	288.80	45.8413	.0025*	(1,4)
		MULTIVARIATE F		
		10.4617	.0886	(3,2)

There is clear evidence of the effect of Latinate diction on graders' judgements; graders overwhelmingly preferred essays written with official-sounding, "mature" vocabulary.

As shown in Table IV earlier, the summed means for Saxon essays were 95.6 (Overall), 92.4 (Mechanics), and 99.6 (Logic), while the summed means for Latinate were 108.4, 103.6, and 107.2 respectively. When the examiners read essays written with Latinate words, they awarded significantly higher grades even on Mechanics and Logic which, presumably, have nothing to do with choice of words. Evidently, the examiners perceived fewer spelling, punctuation and usage errors in Latinate essays, or they reacted with less disapproval of such errors. The examiners also seem to have regarded essays written with longer,

official-sounding words as being better organised and logically more coherent.

The multivariate test just failed to reach statistical significance for differences in diction. This appeared unexpected in view of the fact that all three grading dimensions showed significant differences between Saxon and Latinate essays. The reason was that the variance of Saxon scores over all grading dimensions was close to the variance of all Latinate scores over all grading dimensions, so that differences between them were relatively small compared to the size of differences within the individual grading dimensions. The main differences and standard deviations of differences for all comparisons in 1976 data are included for reference in Appendix D, together with full details of stepdown F ratios from the analyses of variance.

Because the results so far have shown no interaction effect, no differences between Active and Passive, and significant differences between Saxon and Latinate, it follows logically that Active/Saxon essays received significantly lower grades than Passive/Latinate. From Table IV, the means for Style I (A/S) essays were 48.2 (Overall), 47 (Mechanics), and 50.2 (Logic), while for Style IV (P/L) essays the means were 54.8, 51.6 and 54.2 respectively. Styles I and IV represent the extremes in the experimental styles, and, clearly, the examiners much preferred Latinate essays regardless of voice. The importance of what this preference means is realised if the samples below from essays are now inspected, or, indeed, if the complete versions are read at this point from Appendix A.

Essay W:

I Active/Saxon: "The standard of living has rised in Australia and so people have to do less work."

II Active/Latinate: "The standard of living has improven in Australia and therefore individuals are not obliged to labour so diligently."

III Passive/Saxon: "The standard of living in Australia has rised and so less work has to be done by people."

IV Passive/Latinate: "The standard of living in Australia has improven and therefore there is not the necessity for such diligent labour by people as was the case hitherto."

Essay X:

I Active/Saxon: "When a member of parliament who is earning 3 or 4 times the average wage looks upon most people who sit back and be quiet and calls them soft he is judging a book by its cover, he is looking at one side of a case."

II Active/Latinate: "When a politician who receives 3 or 4 times the average wage surveys the majority of citizens who avoid publicity and remain silent and labels them as soft he is forming speculative conclusions on the slightest of evidence, he is conducting a unilateral argument."

III Passive/Saxon: "When most people who sit back and be quiet are looked upon and called soft by a member of parliament whose earnings are 3 or 4 times the average wage this is the judging of a book by its cover, the looking at one side of a case."

IV Passive/Latinate: "When the majority of citizens who avoid publicity and remain silent are surveyed critically and labelled soft by a politician whose salary is 3 or 4 times the average wage this is speculation and judgement on the slightest of evidence, the conduct of a unilateral argument by him."

The ornate, verbose and inflated prose of Styles II and IV has an effect on a reader very much different from the simpler, direct and clearer prose of Styles I and III. It seems unlikely that public examiners in English could have chosen to prefer Style IV more than Style III, and it is astounding that they should have chosen Style IV over Style I. The cumbersome nature of the passive styles did not influence the examiners. They preferred, for example, Style II more than Styles I or III, making no significant distinctions between their grades for I and III. They made no significant distinctions between Style II and IV, again despite the weighty clumsiness of IV. As long as the vocabulary sounded mature they rewarded the essay. It seems that they rewarded Latinate essays regardless of clarity of meaning; in fact, they thought Latinate essays were more logically organised.

The comments that the examiners wrote about each essay show some evidence of the reactions they made. All such comments are included for reference in Appendix B, but some are shown below in order to examine the overwhelming influence of Latinate diction on graders' judgements.

	Styles I or III	Styles II or IV
Examiner 1:		
Essay W:	"The grammar and logic are too weak to ... pass."	"... apart from minor errors of haste, has written accurately."
		"... the essay is fairly sound in organisation and grammar..."
Essay Z:	"... lacks cohesion within the paragraphs and from one paragraph to the next."	"This student has strung his material together quite well ..."

	Styles I or III	Styles II or IV
Examiner 2:		
Essay V:	"... overuse of hyphenated words and phrases shows a lacking in vocabulary..."	"Mature vocabulary ..."
	"Organisation is satisfactory."	"Well organised and structured."
	"Sentences generally tend to be too long leading to clumsy construction..."	
Examiner 5:		
Essay X:	"A poorly structured argument none too well expressed."	"The argument is more clearly expressed than in previous 'models'."
Examiner 6:		
Essay W:	"Logic of ... (example) ... is a bit confusing."	"Occasional lapses in mechanical accuracy which seem odd when vocabulary, sentence development, logic is (sic) taken into account."
	"Peculiarities of syntax and sentence structure."	
Examiner 7:		
Essay W:	"Pretty sound in all respects, save for some awkward and clumsy (careless?) expression and usage in places."	"Intelligent and well-developed essay. Errs on the side of pretentiousness to some degree - actually says less than it seems to say..."
Essay X:	"Usage is very average, but the writer makes a reasonable case - although that case diverges from my reading of the topic."	"Beautifully awkward, formal and impersonal, passive style! Good argument, nearly lost in the verbiage."

	Styles I or III	Styles II or IV
Examiner 8:		
Essay W:	"... communicating effectively despite some errors in mechanics."	"... writes fluently."
	"... expression is satisfactory."	"Articulate and mature."
Essay Z:	"... lacks the clarity of expression and sufficient content of argument ..."	"The writer is articulate and shows keen insight."

The above evidence is qualitative and erratic from grader to grader over some essays, but an inspection of Appendix B will show that the above is typical of the reactions of all examiners to most essays. It certainly shows some of the qualities that the examiners perceived as styles changed. Examiners' judgements and perceptions are obviously influenced by diction.

The award of pass or fail can depend to some degree on the diction the student uses. Five of the eight examiners failed two or more essays in Saxon diction only to pass the same essays when written in Latinate diction. When grading for mechanics, six of the examiners failed two or more essays in Saxon and passed them in Latinate. When grading for logic, four of the examiners failed two or more essays in Saxon, yet passed the same essays in Latinate. Only two graders consistently failed or passed the same essay regardless of stylistic changes, but this was only on the logic dimension. As far as Overall and Mechanics were concerned as grading dimensions, all graders at some time changed their judgements from fail to pass at least once as an essay was encountered in Saxon and then Latinate versions.

Appendix B includes the original letter grades assigned by the eight examiners for all experimental essays.

In summary, the 1976 pilot study found that the eight public examiners:

(1) awarded significantly higher grades to essays written in Latinate diction rather than Saxon on all three grading dimensions;

(2) made no significant distinctions between essays written in active voice and essays in passive voice;

(3) awarded significantly higher grades to essays written in Passive/Latinate rather than Active/Saxon, and likewise for Active/Latinate rather than Passive/Saxon, on the grading dimensions of Overall and Logic; a very strong trend in the same direction was observed when they graded for Mechanics;

(4) showed from their comments that they regarded essays written in Latinate diction to have higher competence in mechanics and logical organisation as well as vocabulary.

The next set of data reported concerns the first set of student-teachers who graded essays in 1977. Full and complete data were received from 51 of the original sample of 60 final-year student-teachers. The gradings were treated in the same way as for the 1976 pilot study. The means of the transformed scores for each essay on each grading dimension are shown in Table VIII below.

TABLE VIII

MEANS OF TRANSFORMED SCORES FOR EACH ESSAY ON EACH GRADING DIMENSION: 1977 DATA

	Style I Active/Saxon			Style II Active/Latinate			Style III Passive/Saxon			Style IV Passive/Latinate		
Essay	O	M	L	O	M	L	O	M	L	O	M	L
V	59	57	59	62	61	61	58	56	58	61	62	61
W	54	53	56	63	60	62	56	54	56	63	60	62
X	55	53	56	59	59	59	55	54	56	59	60	58
Y	48	46	50	50	49	51	49	46	50	50	48	50
Z	51	50	52	57	56	56	51	50	53	57	57	58
Means: O	53.4			58.2			53.8			58		
M		51.8			57			52			57.4	
L			54.6			57.8			54.6			57.8

For ease of inspection, the means for each style are presented as "cell means" in Table IX.

TABLE IX

CELL MEANS OF TRANSFORMED SCORES FOR ESSAYS
BY STYLE ON EACH GRADING DIMENSION : 1977 DATA

		DICTION				Summed Row Means	
		Saxon		Latinate			
VOICE	Active	O	53.4	O	58.2	O	111.6
		M	51.8	M	57	M	108.8
		L	54.6	L	57.8	L	112.4
	Passive	O	53.8	O	58	O	111.8
		M	52	M	57.4	M	109.4
		L	54.6	L	57.8	L	112.4
	Summed Column Means	O	107.2	O	116.2		
		M	103.8	M	114.4		
		L	109.2	L	115.6		

The first hypothesis to be tested was the interaction. The interaction effect over the whole table was found to be not significant, as Table X shows.

TABLE X

ANALYSIS OF VARIANCE TABLE FOR INTERACTION
EFFECTS OF VOICE AND DICTION (1977)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
Overall	1.8	2.25	.2081	(1,4)
Mechanics	.2	.1176	.7489	(1,4)
Logic	0.0	0.0	1.000	(1,4)
		MULTIVARIATE F		
		3.8889	.2113	(3,2)

The insignificance of the interaction effect means that any differences in diction were the same, irrespective of voice; similarly, any differences in voice were the same, irrespective of diction.

The second hypothesis tested was for one of the main effects, that is, the Active against the Passive Voice. It was anticipated that the Passive Voice would receive significantly higher awards than the Active. The summary table follows in Table XI.

TABLE XI

ANALYSIS OF VARIANCE TABLE FOR MAIN EFFECTS OF VOICE (1977)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
Overall	.2	.0909	.7781	(1,4)
Mechanics	1.8	1.3846	.3046	(1,4)
Logic	0.0	0.0	1.000	(1,4)
		MULTIVARIATE F		
		.2578	.8528	(3,2)

Like the 1976 examiners, these graders were not influenced significantly in any direction by the effects of active or passive voice. In fact, the last two tables above have shown that these final-year student-teachers were extremely consistent in awarding the same grades for logic from one voice to the other, since there was no variance at all in their awards. From Table IX it can be seen that the summed means for Active over both types of diction

were 111.6 (Overall), 108.8 (Mechanics) and 112.4 (Logic), while the summed means for Passive were 111.8, 109.4 and 112.4 respectively. The graders were no more responsive to essays written with direct and simple syntax than to essays written with less direct and more cumbersome syntax.

When the main effect of Saxon Diction against Latinate Diction was tested, the 1977 graders produced the same results as the public examiners had in the 1976 pilot study. Latinate diction influenced them significantly so as to award higher grades, as Table XII shows below.

TABLE XII

ANALYSIS OF VARIANCE TABLE FOR MAIN EFFECTS OF DICTION (1977)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
Overall	405.0	15.5769	.0169*	(1,4)
Mechanics	561.8	49.7168	.0022*	(1,4)
Logic	204.8	11.5056	.0275*	(1,4)
		MULTIVARIATE F		
		15.8289	.0601	(3,2)

Once again it is evident that graders preferred essays written with official-sounding, "mature" vocabulary. The size of the differences between the means for the two variables can be inspected in Appendix E, together with details of standard deviations of differences and stepdown F ratios. From Table IX

earlier, it can be seen that the summed means for Saxon essays were 107.2 (Overall), 103.8 (Mechanics) and 109.2 (Logic), while the summed means for Latinate essays were 116.2, 114.4 and 115.6 respectively.

Once again, it is apparent that graders perceived fewer spelling, usage and punctuation errors in Latinate essays, or they reacted to them with less disapproval than when the same errors occurred in Saxon essays. Like the public examiners before them, the student-teachers also regarded Latinate essays as being better organised and logically more coherent.

The multivariate test just failed to reach statistical significance for differences in diction. Once again the reason for this was the similarity of variance over all grading dimensions collectively for Saxon essays compared to Latinate essays.

In view of the fact that there was no interaction effect and that the effect of Diction was significant, it follows logically that there would be a significant difference between grades awarded to Active/Saxon essays and those to Passive/Latinate. From Table IX earlier, it can be seen that the means for Active/Saxon essays were 53.4 (Overall), 51.8 (Mechanics), and 54.6 (Logic), while the means for Passive/Latinate essays were 58, 57.4 and 57.8 respectively. From the same table it can be seen that there were similar differences between Passive/Saxon and Active/Latinate essays.

The student-teachers who were about to take up teaching duties in schools held the same views about writing ability as the public examiners who had been teaching for many years. They rewarded writing that uses long and abstract words; they revealed no preference for writing that has direct and vigorous sentence-structure; they gave significantly higher grades to verbose, weighty, abstract and indirect expression, and considered such writing to contain fewer, or less important, errors in punctuation, usage, spelling and logical organisation. Any essay that was written vigorously, simply, clearly and with everyday words received much lower grades, and was considered to be more faulty in organisation and mechanics. It was more likely to be awarded a fail grade.

Such was the overwhelming influence of Latinate diction.

1978 Results

The third phase of the study used a second set of student-teachers as graders.

From the sample of 57 student-teachers in their final year of training, 52 provided full and complete data. The same statistical treatment as for the two earlier phases of the study was done. In addition, another multivariate analysis of variance was conducted using the original unchanged essays which had been

incorporated into the grading procedures. These allowed for the comparison of grades awarded to original and modified essays to see with which style the grades of original essays were most identified. The results of the latter analysis will come later in this chapter.

The means of the transformed scores for each modified essay on each grading dimension are shown in Table XIII.

TABLE XIII

MEANS OF TRANSFORMED SCORES FOR EACH MODIFIED ESSAY ON EACH GRADING DIMENSION: 1978 DATA

		Style I Active/Saxon			Style II Active/Latinate			Style III Passive/Saxon			Style IV Passive/Latinate		
	Essay	O	M	L	O	M	L	O	M	L	O	M	L
	V	62	62	63	67	65	63	61	59	63	62	62	60
	W	57	54	59	64	58	62	54	53	55	64	61	60
	X	54	54	54	57	57	56	54	53	55	56	55	56
	Y	46	45	48	49	48	50	47	46	48	49	49	49
	Z	49	51	51	56	56	56	50	49	51	55	55	56
	O	53.6			58.6			53.2			57.2		
Means:	M		53.2			56.8			52			56.4	
	L			55			57.4			54.4			56.2

For ease of inspection, the means for each style are presented as "cell means" in Table XIV.

TABLE XIV

CELL MEANS OF TRANSFORMED SCORES FOR MODIFIED ESSAYS BY STYLE ON EACH GRADING DIMENSION: 1978 DATA

		DICTION				Summed Row Means	
		Saxon		Latinate			
VOICE	Active	O	53.6	O	58.6	O	112.2
		M	53.2	M	56.8	M	110.0
		L	55	L	57.4	L	112.4
	Passive	O	53.2	O	57.2	O	110.4
		M	52	M	56.4	M	108.4
		L	54.4	L	56.2	L	110.6
Summed Column Means		O	106.8	O	115.8		
		M	105.2	M	113.2		
		L	109.4	L	113.6		

The first hypothesis tested was the interaction. Table XV below shows that the interaction effect over the whole table was not significant.

TABLE XV

ANALYSIS OF VARIANCE TABLE FOR INTERACTION EFFECTS OF VOICE AND DICTION (1978)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
Overall	5.0	.7692	.4300	(1,4)
Mechanics	3.2	.8649	.4051	(1,4)
Logic	1.8	.5455	.5012	(1,4)
		MULTIVARIATE F		
		1.3393	.4545	(3,2)

As in earlier results described for the other phases of this study, the insignificance of the interaction means that any differences in diction were the same, irrespective of voice, and any differences in voice were the same, irrespective of diction.

The second hypothesis tested was for the main effect of the Active against the Passive Voice. The summary table follows in Table XVI.

TABLE XVI

ANALYSIS OF VARIANCE TABLE FOR MAIN EFFECTS OF VOICE (1978)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
Overall	16.2	2.1039	.2206	(1,4)
Mechanics	12.8	1.0407	.3654	(1,4)
Logic	16.2	2.1039	.2206	(1,4)
		MULTIVARIATE F		
		1.8537	.3693	(3,2)

The 1978 graders were not influenced significantly by active or passive voice. From Table XIV earlier above, it can be seen that the summed means for Active Voice over both types of diction were 112.2 (Overall), 110 (Mechanics), and 112.4 (Logic), while the summed means for Passive were 110.4, 108.4 and 110.6 respectively.

When the main effect of Saxon Diction against Latinate Diction was tested, the 1978 graders produced the same results as the graders before them, with the exception that Diction did not produce a statistically significant difference on the Logic grading dimension, as Table XVII indicates below.

TABLE XVII

ANALYSIS OF VARIANCE TABLE FOR MAIN EFFECTS OF DICTION (1978)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
Overall	405.0	14.2105	.0197*	(1,4)
Mechanics	320.0	30.4762	.0053*	(1,4)
Logic	88.2	3.4319	.1376	(1,4)
		MULTIVARIATE F		
		19.9253	.0482*	(3,2)

It can also be seen that the multivariate F was significant, at the 5% level, for the first time in the study. Once again, the effects of "mature" vocabulary are such that the graders awarded significantly higher grades to Latinate essays than to Saxon essays. The size of the differences between means for Saxon and Latinate may be inspected in Appendix F, together with the standard deviations of differences and the stepdown F ratios.

Referring again to Table XIV, it can be seen that the summed means for Saxon essays were 106.8 (Overall), 105.2 (Mechanics) and 109.4 (Logic), while for Latinate essays, the summed means were 115.8, 113.2 and 113.6 respectively.

Here is further support for the conclusion that graders reward Latinate essays much more than Saxon essays, and that they give significantly higher grades for mechanics, such as spelling, usage, and punctuation. The trend is very strong amongst these graders to consider that essays written with long, less common words show better coherence of argument and logical organisation than those essays in which shorter, direct, concrete and vigorous words are used.

An examination of Table XIV above shows that the means for Active/Saxon essays were 53.6 (Overall), 53.2 (Mechanics) and 55 (Logic), while for Passive/Latinate the means were 57.2, 56.4 and 56.2 respectively. Similar trends are evident in the comparison across the other diagonal cells, between Passive/ Saxon and Active/Latinate essays.

It is clear that the 1978 graders were influenced in their grading in almost identical ways to the 1976 and 1977 graders. It will be more convenient to bring all results together in summary form at this point before the remaining results are examined. The order and numbering of results are the same as for the null hypotheses stated in Chapter IV, page 49.

1. Passive/Latinate essays were graded significantly higher than Active/Saxon essays on all dimensions in all phases of the study. The null hypothesis is rejected by all three sets of results.

2. There was no significant difference between the grades awarded essays in Active Voice and the grades awarded the same essays in Passive Voice. For all three phases of the study, the null hypothesis in regard to Voice is accepted.

3. With regard to Diction, all three sets of graders awarded significantly higher grades to essays written in Latinate Diction compared to essays in Saxon Diction. This occurred with grades for Overall Quality by all years of graders; it occurred with grades for Mechanics by all years of graders; it occurred with grades for Logic by the 1976 and 1977 graders, but not the 1978 graders, though the trend was very strong for the 1978 graders ($p<.1376$). Only the 1978 graders were responsible for significant differences in a multivariate analysis; the higher probability of multivariate differences being due to chance in the other two sets of results was as low as .0886. In short, the null hypothesis is rejected by all three sets of results in regard to Overall Quality and Mechanics, and it is rejected by two sets of results in regard to Logic.

It is clear from the judgements of the 111 graders that a style of writing that incorporates Latinate diction is regarded as superior expression, regardless of other factors such as voice and its associated characteristics of sentence-structure. Graders consider that essays that are written with words of Latinate origins are superior in logical organisation, and that their authors have more command of punctuation, usage and spelling.

The conclusion to be drawn from this is that one style or manner of writing is deemed to be of highest value. There is clear evidence that graders in this study believe that there is a best way to express oneself in writing within the terms of the stylistic alternatives provided by the design of this study. Whether Latinate diction is, as it happens, the one factor that contributes to this "best way of writing" or whether there are other factors that exist beyond the scope of this study, are questions that are beside the point. Graders overwhelmingly preferred any essay written with Latinate words, no matter how poor the argument, how clumsy the sentence-structure, and no matter how tediously grandiose the expression. They held some "model" in mind, possibly, but the effect was to award higher grades to restricted styles of expression. Purposeful expression in any other non-Latinate style received lower grades.

How varied in expression were the essays that the students wrote initially? To answer this question, the results of the multivariate

analyses of variance involving the original unchanged essays may now be examined. It was anticipated that the writing of a student in school would be quite flexible and varied in expression, with a style reflecting the purpose for writing, the audience for whom it is meant, and the personality of the writer himself. It was impossible, however, to anticipate with which of the four experimental styles the original student's essay would be comparable in terms of the awards from graders.

The means of the transformed scores for each original essay are shown in Table XVIII below. Since these scores were the result of the Rasch Model procedures involving all the 1978 data, the transformed scores of the modified essays are also shown, as they were earlier in this chapter (Table XIII).

TABLE XVIII

MEANS OF TRANSFORMED SCORES FOR ALL MODIFIED AND ALL ORIGINAL ESSAYS ON EACH GRADING DIMENSION (1978)

	Style I			Style II			Style III			Style IV			Original		
Essay	O	M	L	O	M	L	O	M	L	O	M	L	O	M	L
V	62	62	63	67	65	63	61	59	63	62	62	60	64	62	63
W	57	54	59	64	58	62	54	53	55	64	61	60	57	54	56
X	54	54	54	57	57	56	54	53	55	56	55	56	56	55	56
Y	46	45	48	49	48	50	47	46	48	49	49	49	48	48	50
Z	49	51	51	56	56	56	50	49	51	55	55	56	51	50	52
Means: O	53.6			58.6			53.2			57.2			55.2		
M		53.2			56.8			52			56.4			53.8	
L			55			57.4			54.4			56.2			55.4

From visual inspection of the above table, it can be seen that the Original essays compare variously with different Style scores. The means for each style and originals are presented in Table XIX.

TABLE XIX

CELL MEANS OF TRANSFORMED SCORES FOR ESSAYS BY MODIFIED AND ORIGINAL STYLES ON EACH GRADING DIMENSION (1978)

		DICTION				ORIGINAL	
		Saxon		Latinate			
VOICE	Active	O	53.6	O	58.6	O	55.2
		M	53.2	M	56.8	M	53.8
		L	55	L	57.4	L	55.4
	Passive	O	53.2	O	57.2		
		M	52	M	56.4		
		L	54.4	L	56.2		

The above table provides a clearer indication of how the original essays compared with their stylistic versions. The difference between any pair of means is equal to the mean of the differences between all five pairs of essays in the particular comparison. The analysis of variance, in effect, took the five original essays and compared them with the four modified versions, style by style, both univariately (on one grading dimension at a time) and multivariately (all grading dimensions collectively). In this way, significant differences in variances were tested.

It will be more convenient to examine the mean differences between pairs of essays in Style-against-Original comparisons as in Table XX before the results of the four analyses of variance are reported.

TABLE XX

MEANS AND STANDARD DEVIATIONS OF DIFFERENCES BETWEEN MODIFIED AND ORIGINAL ESSAYS BY STYLE AND GRADING DIMENSION

	Saxon	Latinate
	(I) AS - ORIGINAL	(II) AL - ORIGINAL
Active	O : M = -1.6 SD = .8944	O : M = 3.4 SD = 2.6077
	M : M = -0.6 SD = 1.5166	M : M = 3.0 SD = 2.2361
	L : M = -0.4 SD = 2.0736	L : M = 2.0 SD = 2.8284
	(III) PS - ORIGINAL	(IV) PL - ORIGINAL
Passive	O : M = -2.0 SD = 1.0000	O : M = 2.0 SD = 3.5355
	M : M = -1.8 SD = .8367	M : M = 2.6 SD = 3.2094
	L : M = -1.0 SD = .7071	L : M = 0.8 SD = 3.1145

It is seen that for Styles I (AS) and III (PS), Original essays were generally rated higher, as the means of the differences were all negative. For Styles II (AL) and IV (PL), Original essays were generally graded lower, as the means of the differences were all positive. Analysis of variance was conducted to compare Originals with each Style in turn. If one or more of the Styles did not differ significantly from the Originals, this would be the best indication of the stylistic characteristics of the original essays in terms of their effect on graders.

The first analysis of variance for Style I against Originals provided the following results as shown in Table XXI.

TABLE XXI

ANALYSIS OF VARIANCE TABLE FOR COMPARISON OF STYLE I (AS) AND ORIGINAL ESSAYS (1978)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
I AS - ORIG. Overall	12.8	16.0	.0162*	(1,4)
I AS - ORIG. Mechanics	1.8	.7826	.4264	(1,4)
I(AS)- ORIG. Logic	.8	.1860	.6885	(1,4)
		MULTIVARIATE F		
		22.1111	.0436*	(3,2)

It can be seen that Style I (AS) essays were considered significantly less competent than the Original essays on the Overall grading dimension and for all grading dimensions collectively. There were no significant differences on Mechanics and Logic, which would indicate that the graders did not find more flaws when an essay was changed from its original to the more direct and simple style of Active/Saxon.

When Style II (AL) versions were compared with Originals, the analysis of variance provided the following results, as shown in Table XXII.

TABLE XXII

ANALYSIS OF VARIANCE TABLE FOR COMPARISON
OF STYLE II (AL) AND ORIGINAL ESSAYS (1978)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
II (AL) - ORIG. Overall	57.8	8.5	.0435*	(1,4)
II (AL) - ORIG. Mechanics	45.0	9.0	.0400*	(1,4)
II (AL) - ORIG. Logic	20.0	2.5	.1891	(1,4)
		MULTIVARIATE F		
		3.8819	.2116	(3,2)

Table XXII shows that Style II (AL) essays were considered significantly more competent on the Overall and Mechanics grading dimensions, and the trend is strong but not significant for higher grades to be given to Style II (AL) versions for Logic. It seems a peculiar characteristic of the 1978 graders that they were less inclined than the earlier graders in this study to change their awards for logical organisation of the modified essays (see also Table XVII). Nevertheless, it appears that the original essays were graded significantly lower than Style II (AL) either because the originals lacked the effects of the active voice or because they did not have a preponderance of Latinate vocabulary. Since the originals were generally rated higher than Style I (AS) (Table XXI), but lower than Style II (AL), it would appear that they possessed enough Latinate diction to be regarded superior to Style I (AS) but not enough to be rated equally with Style II (AL). The voice was active in both comparisons so far, so Diction again seemed to be the differential factor.

Table XXIII shows the summary table for the comparison of original essays with the Style III (Passive/Saxon) versions.

TABLE XXIII

ANALYSIS OF VARIANCE TABLE FOR COMPARISON OF STYLE III (PS) AND ORIGINAL ESSAYS (1978)

	MEAN SQUARE	UNIVARIATE F	p<	df
III PS - ORIG. Overall	20.0	20.0	.0111*	(1,4)
III PS - ORIG. Mechanics	16.2	23.1429	.0086*	(1,4)
III (PS)- ORIG. Logic	5.0	10.0	.0342*	(1,4)
		MULTIVARIATE F		
		24.3333	.0398*	(3,2)

Clearly, the graders awarded significantly higher grades to the Originals, and did so consistently over all grading dimensions.

At this stage of analysis, Originals have been seen to be superior to Style I (AS) and Style III (PS), but inferior to Style II (AL). In the light of earlier results showing no Voice effect, Diction would be the operating variable in these comparisons. Apparently, Originals possessed enough Latinate diction for them to be rated significantly higher than either of the two Saxon styles. But because the Style II (AL) versions were so uniformly and intensely weighted with Latinate diction, it seems evident that the Originals

could not "match" them and were, consequently, rated significantly lower than Style II (AL). Conversely, Originals would be rated significantly higher than any Saxon essays, regardless of voice, because Originals contained some Latinate diction at least.

The results of the final comparison between Originals and Style IV (PL) are shown below in Table XXIV.

TABLE XXIV

ANALYSIS OF VARIANCE TABLE FOR COMPARISON
OF STYLE IV (PL) AND ORIGINAL ESSAYS (1978)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	df
IV PL - ORIG. Overall	20.0	1.6	.2746	(1,4)
IV PL - ORIG. Mechanics	33.8	3.2816	.1443	(1,4)
IV (PL)- ORIG. Logic	3.2	.3299	.5965	(1,4)
		MULTIVARIATE F		
		1.8938	.3640	(3,2)

There was no statistically significant difference between grades awarded Originals and Style IV (PL) essays. This result is surprising in the light of earlier results which indicate no Voice effect and the statistical equivalence of essays that are Latinate. It is consistent with the other comparisons to expect Originals to have been graded significantly lower than Style IV (PL) versions.

An examination of Table XX provides the most likely explanation (p 93). It can be seen that the means of differences for Style IV (PL) scores minus Original scores are 2 (Overall), 2.6 (Mechanics) and 0.8 (Logic). These differences are amongst the highest in the whole of Table XX, with the exception of the differences for Logic. But these differences fail to reach statistical significance. The standard deviations of the differences show a wide variance in the graders' reactions, and it is the size of these variances which would have prevented the F ratio from reaching significance at the 5% level. It seems evident that the differences in variance tested by the *F* ratio are too small in relation to the wide error variance, whereas, elsewhere from Table XX, small differences in variance showed as being significant because of low error variance.

It is evident that graders varied widely in awarding different grades to Style IV (PL) and its original essay. It is reasonable to expect that some 71% of the graders would have awarded the Style IV version the same grade or a higher grade than the Original; this is based on the mean and standard deviation of differences on the Overall grading dimension and based on the assumption of a fairly normal distribution of differences. Similarly, some 79% of graders would have awarded Style IV the same or higher grade than the Original essay on the Mechanics dimension.

In the light of the strong and consistent trends in the data reported, it can be concluded, as before, that the writing of students which was varied, natural and flexible was not regarded by essay-graders in this study to be of the highest quality. The "best" writing is that which is indirect, official-sounding, weighty and protracted in style.

The essays that will be rewarded with much higher grades are those that use long and uncommon words of Latinate origins, even if those essays are mediocre in spelling, punctuation, usage and logical organisation.

CHAPTER VI

SUMMARY, CONCLUSIONS AND IMPLICATIONS

This study has investigated the effects on essay-graders when stylistic devices have been manipulated in the essays to be graded. Five essays written by final-year secondary students were each re-written in four versions so as to build in stylistic changes. These changes involved the use of the active and passive voice in combination with words of Anglo-Saxon and Latinate/Greek origins.

It was hypothesised that:

(1) essay-graders would award higher grades to essays in the Passive/Latinate style compared to the same essays in Active/Saxon style;

(2) essay-graders would award higher grades to essays in the passive voice compared to the same essays in the active voice;

(3) essay-graders would award higher grades to essays using Latinate diction compared to the same essays using Saxon diction.

Comparisons were also made of the grades awarded to original, unchanged essays with the grades awarded to each of the four stylistic versions.

Eight public-examiners in English and one hundred and three final-year student-teachers graded the essays on a five-point scale, awarding grades on three grading dimensions, namely, Overall Quality, Mechanics, and Logic and Organisation. The experiment was conducted over three years, and the data were treated by multivariate analysis of variance using a repeated-measures design. The actual scores used in the analyses of variance were transformations from the graders' initial grade-scores. These transformations were computed using the Rasch Polychotomous Rating Model which allowed for more precise measures of the students' writing ability to be used.

The results were very consistent over all three phases of the study. The conclusions that are stated below are clearly and consistently supported by the results of all analyses of the data.

(1) Graders give significantly higher grades to an essay when it is written in a Passive/Latinate style than when it is written in an Active/Saxon style. Similarly, when an essay is Active/Latinate in style, it is rewarded with significantly higher grades than when it is written in Passive/Saxon style.

(2) Graders give the same grades to essays written in the active and the passive voice.

(3) Graders award significantly higher grades to an essay when it is written with Latinate diction rather than Saxon diction. They give significantly higher grades to Latinate essays, regardless of voice, and these higher grades are awarded not only for overall quality, but also for the standard of spelling, punctuation, usage, logic and organisation.

(4) Graders are so influenced by the use of Latinate diction that the more a student incorporates Latinate words in his essay, the higher his grade will be.

(5) Graders are so influenced by the use of Latinate diction in essays that they overlook the errors in punctuation and spelling, usage and logic, and award significantly higher grades on these dimensions. Their perceptions of an essay change radically as the essay is modified from Saxon to Latinate diction, even though a change in diction can in no way alter the logic of an argument.

(6) As a result of the above, graders prefer the more ornate, grandiose, cumbersome, abstract and indirect use of words and sentences significantly more than the simple, lucid, vigorous and direct use of words and sentences. Graders

reward official-sounding, lengthy and convoluted writing significantly more than they do clear and unpretentious writing.

(7) When a student writes with his own choice of a variety of words and flexibility in voice, his essay will be graded significantly higher than if he wrote using Saxon words only. However, his essay will be graded significantly lower than if he had saturated his essay with Latinate words, no matter how pretentious, artificial, verbose and ponderous. Again, the Latinate essay will be regarded as being far better organised logically, even though the ideas, material and handling of the argument are exactly the same.

(8) Both experienced and inexperienced teachers are likely to react in the ways described above.

The implications arising from the results and conclusions of this study are disturbing. It is not likely that students in schools are aware that simple, clear and direct writing is not valued as highly as more complex, abstract and flaccid writing. If they are being told to write simply, vigorously, and never to use a long word when a short one will do, they are being deceived. On the contrary, their writing will be rewarded most when they learn to ignore such advice. If, however, they are instructed by teachers to use long, grandiose and abstract words, and never to use a short

word when they can think of a longer one, then the pragmatic purpose of obtaining higher grades for written expression will probably be achieved at the expense of writing that is creative, flexible and honest.

It is disturbing that examiners and teachers should be less receptive to the coherence and cogency of an essay than to the ill-conceived notion of "mature vocabulary". Can logic be regarded so insensibly by teachers and examiners that it is believed to improve if big words are used? The implication that follows here is that teachers have to re-appraise what they think the elements of effective communication are, and the whole notion of "maturity" of expression. Until this re-appraisal and retraining occur, teachers will continue to assess writing ability invalidly.

Long words and the use of the passive voice might be highly approved because of the social prestige that is ascribed such expression. But the implication from this is frightening, for to place higher value on officialese that attracts social prestige and to devalue that expression which is clear and direct is to prostitute the language in ways that Orwell predicted. The fact that schools could be encouraging, knowingly or unknowingly, the development of children's writing along these lines is frightening indeed. But the evidence from the one hundred and eleven graders in this study, and the evidence cited from American studies and classroom practices, shows that children are being subjected to

these very effects - at least by way of the assessment of their writing, and, in some cases, by intentional instruction.

There is some cause for concern that even the informed and enlightened teachers who teach English by the "growth" model may be evaluating their children's writing in ways contrary to their philosophy. New strategies for English teaching must be accompanied by revised and compatible criteria for evaluation. If the results of this study suggest that teachers old and new in the schools are evaluating essays in inconsistent ways, then it is clear that the shift towards new objectives and curricula in English have not been accompanied by an appropriate shift in evaluation criteria. It seems that the questions raised in Chapter I have, in fact, disturbing answers, since it is very likely that many teachers are holding onto views of essay-writing evaluation that are now out-moded when compared to new curricula and objectives in English.

Each piece of research that throws light on the validity or otherwise of evaluation procedures will always invite statements that indict the persons concerned. It is so easy to use hindsight to criticise inefficiencies in educational practices. In fairness to teachers and examiners of English, it must be said that they cannot be indicted for performing in certain ways when the inadequacies of those ways are unknown. But when specific

inadequacies are brought to light, as has been done in this dissertation, the immediate implication that matters is that such faults be corrected. In this regard, awareness by teachers and examiners of their past practices comes first, and correction through re-appraisal and retraining must follow.

BIBLIOGRAPHY

BIBLIOGRAPHY

Andrich, D. Latent trait psychometric theory in the measurement and evaluation of essay-writing ability. Ph.D. dissertation, University of Chicago, 1973.

Andrich, D. "The Rasch multiplicative binominal model: applications to attitude data". Research Report No. 1. Measurement and Statistics Laboratory, Department of Education, University of Western Australia, 1975.

Andrich, D. "Scaling attitude items constructed and scored in the Likert tradition". Educational and Psychological Measurement, 1978a 38 (3), 665-680.

Andrich, D. "Application of a psychometric rating model to ordered categories which are scored with successive integers". Applied Psychological Measurement, 1978b, 2 (4), 581-594.

Barnes, D. From communication to curriculum. Harmondsworth: Penguin, 1975.

Barnes, D., Britton, J., Rosen, H., & the L.A.T.E. Language, the learner and the school. Harmondsworth: Penguin, 1971.

Bennett, B. "Learning to write". English in Australia, 1979, 50, 5-16.

Bennett, B. H., & Hay, J. A. Directions in Australian secondary school English. Camberwell, Victoria: Longmans, 1971.

Bennett, B., & Walker, A. "The development of writing abilities, 15 to 17: Research in Progress". English in Australia, 1978, 44, 5-15.

Bernstein, B. Class, codes and control, Vol. 1. London: Routledge and Kegan Paul, 1971.

Boulton, M. The anatomy of prose. London: Routledge and Kegan Paul, 1954.

Britton, J. Language and learning. Harmondsworth, Penguin, 1970.

_____ "Language and the nature of learning: An individual perspective". In 76th Yearbook of N.S.S.E., The teaching of English. Chicago: N.S.S.E., 1977, 1-38.

Bruner, J. S. The relevance of education. London: George Allen and Unwin, 1972.

Burgess, C., et al. Understanding children writing. Harmondsworth: Penguin, 1973.

Cast, B. M. P. "The efficacy of different methods of marking English composition". British Journal of Educational Psychology, 1939, 9, 256-269.

Chomsky, N. Syntactic structures. The Hague: Mounton, 1957.

_____ Language and mind (2nd ed.). New York: Harcourt Brace Jovanovich, 1972.

Creber, J. W. P. Lost for words. Harmondsworth: Penguin, 1972.

Devine, T. G. "What is English?" In F. L. Steeves (Ed.). The subjects in the curriculum. New York: Odyssey Press, 1968.

Dewey, J. Democracy and education (1916). New York: Macmillan, 1961.

Diamond, C. T. P. "The skills or growth: English teachers' thinking about the teaching of writing". English in Australia, 1979, 48, 57-66.

Dixon, J. Growth through English. London: Oxford University Press, 1975.

Eltis, K. J. "Pupils' speech-style and teacher reaction: Implications from some Australian data". English in Australia, 1980, 51, 27-35.

Gosling, G. W. H. Marking English compositions. Hawthorn, Victoria: A.C.E.R., 1966.

Guth, H. P. English for a new generation. New York: McGraw-Hill, 1973.

Hake, R. Composition theory in identifying and evaluating essay-writing ability. Ph.D. dissertation, University of Chicago, 1973.

Hake, R., & Williams, J. M. "An investigation into overt and tacit values used by composition teachers: What do writing teachers teach?" (Mimeograph) University of Chicago, 1975.

_____ "Style and its consequences". (Mimeograph) University of Chicago, 1979. (Also in press).

Halliday, M. A. K. "The language acquisition and development project". In Focus on Language; Occasional Paper of the National Committee on English Teaching. Canberra: Curriculum Development Centre, 1977a.

_____ "How children learn language". In K. D. Watson, & R. D. Eagleson (Eds.). English in secondary schools: Today and tomorrow. Ashfield, N.S.W.,: English Teachers' Association of N.S.W., 1977b.

Hartog, Sir Phillip. The marking of English essays. London: Macmillan, 1941.

Hartog, Sir Phillip, & Rhodes, E. C. The marks of examiners. London: Macmillan, 1936.

Kelly, G. A. A theory of personality. New York: Norton, 1963.

Kerlinger, F. N. Foundations of behavioural research, (2nd ed.). New York: Holt, Rinehart and Winston, 1973.

King, A. Writing. London: Longmans Green & Co., 1955.

Labov, W. "The logic of non-standard English". In F. Williams (Ed.), Language and poverty. Chicago: Markham, 1970.

Lawton, D. "Social background and the teaching of English". In J. Britton (Ed.). Handbook for English teachers, No. 2. Talking and writing. London: Methuen, 1966.

Lawton, D. Social class, language and education. London: Routledge and Kegan Paul, 1968.

Martin, N. "What are they up to?" In A. Jones, & J. Mulford (Eds.). Children using language. London: Oxford University Press, 1971.

Martin, N. et al. Understanding children talking. Harmondsworth: Penguin, 1976a.

_____ Writing and learning across the curriculum 11-16. London: Ward Lock Educational, 1976b.

Murray, D. M. A writer teaches writing. Boston: Houghton Mifflin, 1968.

Moffett, J. Teaching the universe of discourse. Boston: Houghton Mifflin, 1968.

Moffett, J., & Wagner, R. J. Student-centred Language Arts and Reading, K-13, (2nd ed.). Boston: Houghton Mifflin, 1976.

Orwell, G. "Politics and the English language". Horizon, April, 1946, 76.

Rasch, G. Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research, 1960.

_____ "On general laws and the meaning of measurement in psychology". In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, California: University of California Press, 1961.

_____ "An individualistic approach to item analysis". In Lazarsfield, & Henry (Eds.). Readings in mathematical social science. Chicago: S.R.A., 1966, 89-107.

Robinson, W. P. "Language in and out of school". English in Australia, 1976, 34, 14-25.

Rosen, C., & Rosen, H. The language of primary school children. Harmondsworth: Penguin, 1973.

Shoenheimer, H. P. (Ed.). English in Australian secondary schools. Melbourne: Cheshire, 1972.

Scott, A. F. Meaning and Style, (2nd ed.). London: Macmillan, 1942.

Strunk, W., & White, E. B. The elements of style. New York: Macmillan, 1972.

Thomson, G. H. The factorial analysis of human ability. London: University of London Press, 1939.

Warner, A. A short guide to English style. London: Oxford University Press, 1961.

Wiley, D. E. "Latent partition analysis". Psychometrika, 1967, 32 (2), 182-194.

Wiseman, S. "The marking of English composition in grammar school selection". British Journal of Educational Psychology, 1949, 19, 200-209.

Witte, S. P., & Davis, A. S. "The stability of T-Unit length: A preliminary investigation". Research in the Teaching of English, 1980, 14 (1), 15-17.

Wright, B. D., & Stone, M. H. Best test design. Chicago: Mesa Press, 1979.

APPENDICES

APPENDIX A

NOTES ON THE RE-WRITING OF ESSAYS AND SAMPLES OF STYLISTIC VERSIONS

APPENDIX A

NOTES ON THE RE-WRITING OF ESSAYS AND SAMPLES

OF STYLISTIC VERSIONS

The re-writing involved two techniques in the main:

(a) inverting sentences between loose and periodic forms in order to produce the required voice in association with intransitive verb forms

(b) changing most key words in every sentence to make a consistent vocabulary drawn from Latin or Greek roots, or drawn from Anglo-Saxon, Scandinavian or even Germanic roots.

There were some problems associated with strict etymological derivation and retaining the same meaning as much as possible. The problem was solved as needed by using a common and simple word for a referent and the "official-sounding" word for the same referent. In other cases, where derivation of the word was in doubt or diffuse, the word was nominalised as an "official-sounding" noun or a simple verb in accordance with the voice parameter being used. The effect was to produce a version of the experimental style which was consistent in its impact on a reader.

Following are examples of the four versions and their "original". Only three of the five essays have been appended. Two were excluded only in the interests of relative brevity.

Life is becoming too easy for the typical Australian. The old days of honor, strictness, hardworking and discipline are being ovetaken by a permisive vulnerable society who are relying on technological equipment to do their work for them.

As many parents have been heard to say 'I wasn't allowed to do that when I was a child' this repeats generation after generation. Parents are becoming too soft on their children, they are giving in to childrens wishes and consequently the majority of Australians are becoming soft. Children are allowed to use dangerous machines eg young boys on minibikes, girls are allowed to go out at early ages and are also becoming promiscuous, crimes committed by children and adults are being delt with more leniently and excuses are being made for dangarous actions of people. Marriage ages have dropped, divorce comes easily, unmarried mothers are accepted. The morals of Australian people are becoming more broadminded and consequently crimes and violence are increasing, people don't bother to get married because they live together anyway, drugs and alcohol are becoming widely used and there is a general breakdown of the family group.

The standard of living of Australia has rised and therefore people have to do less work. This is truly bought out in the use of computers who do mans thinking for them. Australians are eating rich foods in plentiful supply and are beginning to suffer because of it (heart attacks from obesity). Alcohol is being consumed in enormous quantities and is creating 'the great Australian gut' in many males. As well as food coming to easily to the Australians, automobiles have made them lazy to the point where a car is essential and walking is only practised from the parking area to the office. The housewife has modern technological aids which make housework easy. In general Australians are relying on their high standard of living and its benefits instead of standing on their own feet and looking after themselves and others. They are being pampered to an extent where luxury goods are considered necessary.

BEST COPY AVAILABLE

From this comes people who go on the dole and don't do any work. The genuine reasons for this benefit are fast becoming minor as cheats try to sponge of it. Australians are taking advantage of social benefits which the government provides. This not only happens within the lower and middle class workers but also people such as doctors are suspected of cheating medibank. Where is the ancient honor which existed between our ancestors? People worked for a living about one hundred years ago. They made their lives themselves and were independent in contrast to a number of Australians who have worked for only a short period of their lives and have retired to enjoy the gay social life. The values that Australians place on things now have become very materialistic and they are changing the peace and tranquillity into an unfeeling group of people. Respect is lost. Australians are becoming too soft.

Allowances are made for people's actions, the younger generation have been getting essential things of life because their parents may have done well and can provide their children with luxuries. But, what good does this do the children? They don't place high values on things like cars, clothes because they haven't had to work for it. They were given it. The younger generation of Australians are having everything 'handed to them off a plate'. As a result of this the following generation is given things which is developing selfishness and a group of people who have no values.

We can take this to be a change from the native aborigine who cared for himself and his family, worked hard, had honor and respect for his forefathers to the permissive, lazy, pampered people who cover the majority of the Australian people. Of course there are still Australias who haven't let high standards of living effect them to such an extent, however they are fast becoming a minority.

As a consequence of the vulnerability of Australians crimes are increasing, penalties are becoming lenient, people are doing exactly what they like and are gradually becoming uncontrollable. Australians are indeed becoming too soft. Technology has been a source of 'Fatal Impact' (by Alan Moorehead) on the Australian people, causing their general breakdown of morals and attitudes and a rise of the age of materialism.

ESSAY W: STYLE I (ACTIVE/SAXON)

The average Australian gets life too easy. Today's slack unguarded society relies on the latest gear to do their work for them and is ovetaking the old days of honor, firmness, hardworking and sticking to the rules.

As many parents say 'I wasn't allowed to do that when I was a child' this carries on year after year. Parents are becoming too soft on their children, they are giving into childrens wishes and so most Australians are becoming soft. Children use dangerous machines eg young boys on minibikes, parents allow girls to go out at early ages and girls are becoming sexually sinful, people are soft on crimes that children and grown-ups do and they make excuses when people act dangarously. People marry at earlier ages, they divorce easily, people don't worry about unwedded girls who are having a baby. Australian people are getting more broadminded in their morals and so crimes and roughness are growing, people don't bother to marry because they live together anyway, people everywhere drink and use drugs more and the family is breaking down.

The standard of living has rised in Australia and so people have to do less work. The use of computers has truly bought this out, computers who do mans thinking for them. Australians are eating large doses of rich foods and are starting to suffer because of it (heart attacks from fatness). They drink alcohol in huge amounts and are bringing about 'the great Australian gut' in many men. As well as getting food to easily, the Australians are lazy from using cars to the point where they must always have a car and they only walk from the parking lot to the office. The housewife finds housework easy with the latest technical aids. Over all, Australians rely on their high standard of living and its outcomes instead of standing on their own feet and looking after themselves and others. They spoil themselves to the point where they think that they must have luxury goods.

From this comes people who go on the dole and don't do any work. The good reasons for this pay-out are fast becoming smaller as cheats try to sponge of it. Australians are over-working the social pay-outs which the government gives. Not only lower and middle class workers do this but also we think people such as doctors are cheating medibank. Where is the old honor which held our forefathers together? People worked for a living about one hundred years ago. They made their own lives and stood up for themselves as against some Australians who have worked for only a short period of their lives and have stopped work to live it up. Australians now place very materialistic values on things and they are changing the peace and quietness into an unfeeling group of people. We have lost all respect. Australians are becoming too soft.

We have to make allowances for what people have done, the younger people have been getting essential things of life because their parents may have done well and can give their children the best of everything. But, what good does this do the children? They don't place high values on things like cars, clothes because they haven't had to work for it. They just take it. We hand everything to the younger Australians 'off a plate'. As a result the young people following us take all we give and they are a selfish group of people who value nothing.

We are a change from the native aborigine who cared for himself and his family, worked hard, had honor and respect for his forefathers to the easy-going, lazy, spoilt people who cover most of the Australian people. Of course there are still Australias who haven't let high standards of living effect them to such an extent, but they are fast becoming fewer.

Because Australians do not know how much they are slipping they are doing more crimes, punishing more softly, people are doing exactly what they like and are little by little getting out of hand. Australians are indeed softening themselves up. Technology has made a 'Fatal Impact' (by Alan Moorehead) on the Australian people, causing them to lose hold of their morals and attitudes and to begin the age when everyone grabs what he can get.

ESSAY W: STYLE II (ACTIVE/LATINATE)

The typical Australian finds life excessively comfortable. Contemporary permissive and vulnerable society depends on technological equipment to perform their labours for them and is usuping the old days of honor, strictness, toiling and discipline.

As numerous parents remark 'I was not permitted to do that when I was a child' this repeats generation after generation. Parents are increasingly lenient in the extreme with their children, they are surrendering to childrens demands and consequently the majority of Australians are becoming soft. Children are manipulating hazardous machines eg youthful males on minibikes, parents permit girls to socialise publicly at immature ages and girls are furthermore becoming promiscuous, the community treats more leniently crimes that juveniles and adults commit and people fabricate rationalisations when individuals behave dangarously. Increasingly younger individuals enter matrimony, they annul their marriage expeditiously, the community accepts unmarried females who are pregnant. Australian people increasingly are developing a less myopic view of their ethics and consequently crimes and violence are increasing, individuals avoid exerting themselves to enter matrimony because they cohabit in any event, people from various and several sectors of the community consume narcotics and alcohol extensively and the family group is generally collapsing.

The standard of living has improven in Australia and therefore individuals are not obliged to labour so diligently. The deployment of computers has genuinely bought this out, computers who perform mans reasoning for them. Australians are consuming elaborate diets in plentiful supply and are now reaping the consequences of it (cardio-vascular arrest from obesity). They consume intoxicating liquor in enormous quanitities and are creating 'the prominent Australian paunch' in numerous males. In addition to receiving food to easily, the Australians are indolent as a consequence of utilising automobiles to such an extent that they consider it imperative to possess a vehicle and they restrict their practice of

walking to that from the parking area to the office. The wife at home finds homeduties painless with modern technological implements. In general, Australians are depending on their high standard of living and its benefits instead of being independent and nurturing themselves and others. They pamper themselves to the extent where they consider it imperative to possess luxury items.

From this arises persons who live on social security and remain unemployed. The genuine justifications for this benefit are rapidly diminishing as frauds endeavour to sustain their lives entirely upon it. Australians are taking advantage of social benefits which the government provides. Not only plebeian and bourgeois employees act in this mode but in addition we suspect that persons such as physicians are indulging in deceptive practices against medibank. Where is the ancient honor that existed between our ancestors? The populace laboured to survive about a century ago. They nurtured their existence themselves and were independent in contrast to a number of Australians who have employed themselves for only a temporary period of their existence and have retired to enjoy the gay social life. Australians now attribute extremely materialistic values to items and they are transforming the serenity and tranquillity into a group of people who have no emotions. We have sacrificed all respect. Australians are becoming too soft.

We have to make concessions for what people have done, the younger generation have been acquiring the necessities of life because their parents may have prospered and can provide their children with luxuries. However, what benefit do the children derive from this? They fail to appreciate the value of commodities such as automobiles, wearing apparel because they have not been compelled to labour towards the possession of it. We provided it. A life of luxuriousness we hand to the younger generation 'off a platter'. As a consequence of this we readily

provide commodities to the ensuing generation which is developing them into an ego-centric body of individuals who possess no values.

We must admit that we are a radical departure from the indigenous aborigine who nurtured himself and his dependants, toiled strenuously, maintained honor and esteem for his ancestors to the permissive, slothful, pampered people who cover the majority of the Australian nation. Obviously there exist yet Australias who have not succumbed to the degenerating affects of superior standards of living to any comparable degree, however they are rapidly developing into a minority.

As a consequence of the vulnerability of Australians they are committing more crimes, imposing penalties with increasing leniency, individuals behave precisely as they desire and are gradually becoming uncontrollable. Australians are indeed making themselves too soft. Technology has wrought a 'Fatal Impact' (by Alan Moorehead) on the Australian nation, causing them to suffer general collapse of their ethics and attitudes and to launch into the age of materialism.

ESSAY W: STYLE III (PASSIVE/SAXON)

Life is getting too easy for the average Australian. The old days of honor, firmness, hardworking and sticking to the rules are being ovetaken by a slack unguarded society who are relying on the latest gear to have their work done for them.

As many parents have been heard to say 'I wasn't allowed to do that when I was a child' this is repeated year after year. Children are becoming too softly treated by their parents, their wishes are given into by parents and so most Australians are becoming soft. Children are allowed to use dangerous machines eg young boys on minibikes, girls are allowed to go out at early ages and are also becoming sexually sinful, crimes done by children and grown-ups are being dealt with more softly and excuses are being made for dangarous actions of people. There has been a drop in marriage ages, divorce is done easily, unwedded girls who are about to have a baby are regarded as being all right. There is more broadmindedness in the morals of Australian people and so there is a growth in crimes and roughness, people are not bothered about getting married because they are in the position of living together anyway, drugs and drink are becoming widely used and there is an over-all breakdown of the family.

The standard of living in Australia has rised and so less work has to be done by people. This is truly bought out in the use of computers who do mans thinking for them. Large stocks of rich foods are eaten by Australians who are now in the position of beginning to suffer because of it (heart attacks from fatness). Alcohol is being drunk in huge lots and 'the great Australian gut' in many men is brought about by this. As well as food coming to easily to the Australians, they have been made lazy by the use of cars to the point where a car is always needed and walking is only done from the parking lot to the office. The housewife has modern technical aids by which housework is made easy. Over all, the high standard of living and its outcomes are being relied on by Australians instead

of standing on their own feet and looking after themselves and others. They are being spoilt to the point where only the best is thought to be a must.

From this comes people on the dole and not doing any work. There are fewer and fewer good reasons left for this pay-out as it is being sponged of by cheats. The social pay-outs given by the government are being over-worked by Australians. This not only happens within the lower and middleclass workers but also people such as doctors are thought to be cheating medibank. Where is the old honor which was held by our forefathers? A living was worked for about one hundred years ago. Lives were made by people standing up alone as against some Australians whose work has been only for a short period of their lives and who are finished with work now to live it up. The values that are placed on things by Australians now are very materialistic and the changing of the peace and quietness into an unfeeling group of people is being done by them. Respect is lost. Australians are becoming too soft.

Allowances are made for people's actions, the essential things of life have been got by the younger people because their parents may be in the position of having done well and their children can be given the best of everything. But, what good is this to the children? Things like cars, clothes are not highly thought of by them because no work has had to be done for it. They were given it. The younger Australians are having everything 'handed to them off a plate'. As a result of this the young people following us are given things from which is brought about selfishness and a group of people with no values.

There has been a change from the native aborigine who is seen to have cared for himself and his family, to have worked hard, while having honor and respect for his forefathers to the easy-going, lazy, spoilt people who cover most of the Australian people. Of course there are still Australias who have not been effected to such an extent by high standards of living, but there are fewer of them.

Because of the slackness and carelessness of Australians there is a growth in crimes, punishment is softer, there is the fact that exactly what is wanted to be done by people is done by them and they are little by little getting out of hand. Australians have indeed become too soft. A 'Fatal Impact' (by Alan Moorehead) has been made on the Austraian people by technology, and a wide breakdown of their morals and ideas together with a new age of everything being grabbed by everyone has been caused.

ESSAY W: STYLE IV (PASSIVE/LATINATE)

Life is becoming excessively comfortable for the typical Australian. The old days of honor, strictness, toil and discipline are being usuped by a permissive vulnerable society who are dependent upon technological equipment to have their labours performed for them.

As numerous parents have been observed to remark 'I was not permitted to do that when I was a child' this is repeated generation after generation. Increasingly children are in the situation of being treated with extreme lenience by their guardians, their demands are surrendered to by parents and consequently the majority of Australians are becoming soft. Children are permitted to manipulate hazardous machines eg youthful males on minibikes, girls are permitted to socialise publicly at immature ages and are furthermore becoming promiscuous, crimes committed by juveniles and adults are being treated with greater leniency and rationalisations are fabricated for dangarous behaviour of individuals. There has been a substantial decline in ages of entry into matrimony, marriage annulment is an expeditious procedure, unmarried mothers are accepted. Australian people are increasingly more accommodating in their ethical views and consequently there exists an increase in crimes and violence, individuals are not concerned about matrimony since they are in cohabitation in any event, narcotics and alcohol are consumed extensively throughout the community and there is a general collapse of the family unit.

The standard of living in Australia has improven and therefore there is not the necessity for such diligent labour by people as was the case hitherto. This is genuinely bought to notice in the deployment of computers who perform mans reasoning for them. Elaborate diets in abundant supply are being consumed by Australians whose suffering as a consequence of it has been commenced (cardio-vascular arrest from obesity). Intoxicating liquor is being consumed

in enormous quantities and the creation in numerous males of 'the prominent Australian paunch' is the result. In addition to food being to readily available to the Australians, they have been made indolent by the utilisation of automobiles to such an extent that the possession of a vehicle is considered imperative and walking is only practised from the parking area to the office. The wife at home has modern technological implements by which homeduties may be performed with facility. In general there is a dependence by Australians upon their elevated standard of living and its benefits instead of upon independence and the nurture of self and others. They are being pampered to the extent where luxury items are considered essential.

From this arises persons continuously unemployed and provided for by social security. The genuine justifications for these benefits are rapidly becoming minor as they are fraudulently exploited as a means of survival. The social benefits provided by the government are being exploited by Australians. This occurs not only within the plebeian and bourgeois strata of society but also persons such as physicians are suspected of indulging in deceptive practices against Medibank. Where is the ancient honor which existed between our ancestors? Survivial was wrought by toil a century ago. The livelihood of persons was fashioned individually and independently in contast to a number of Australians for whom only a temporary measure of their existence has been devoted to gainful employment and who are presently in a state of self-imposed retirement in order to relish an exuberating social life. The values that are attributed by Australians to various items are now extremely materialistic and the transformation of the serenity and tranquillity into a body of callous people is being wrought by them.

Concessions are granted for the actions of some persons, the necessities of life have been acquired by the younger generation because of the well-deserved prosperity of their parents whose children can be provided with luxuries. However, what benefit accrues from this to the children? Commodities exemplified by

motor vehicles, wearing apparel are not valued by them because there has been no exertion of effort for it. They were provided it. The younger generation of Australians are having a life of luxury 'handed to them off a silver platter'. As an inevitable consequence of this, the ensuing generation is readily provided with commodities from which condition is generated an ego-centric body of individuals with a value-system of negligible proportions.

A radical departure must be conceded here from the indigenous aborigine who is perceived to have nurtured himself and his dependants, to have toiled strenuously, while maintaining honor and esteem for his ancestors to the permissive, slothful, pampered individuals who cover the majority of the Australian nation. Obviously there exist yet Australias who have not been contaminated to any comparable degree by the degenerating affects of superior standards of living, however theirs is a rapid decline towards the status of a minority.

As a consequence of the vulnerability of Australians there is an increase in the frequency of crimes, penalties are increasingly lenient, it is evident that behaviour precisely according to whatever is fancied is actually performed by individuals who are gradually becoming uncontrollable. Australians have indeed become too soft. A 'Fatal Impact' (by Alan Moorehead) has been wrought on the Australian nation by technology, and a general collapse of ethics and attitudes together with the emergence of the age of materialism is the consequence.

ESSAY X: ORIGINAL

It was said by a politician recently that "Australians are Becoming Soft". This brings to mind an idea, Is the Australian image softening along with its people? The most common image of an Australian is a beer guzzling, quick tounged person. This image is surely a minority of the population and therefore a "popular" image should not be based on this alone.

Today's Australian is basically an urbanite who spends most of his time withing an area arround his town or city. Therefore the old Australian image of the "Man on the land" is also negated.

So we have a picture of the modern Australian image. An urbanite who lives and works within the city during the week and relaxes at home on the weekend. With his pay he may be purchasing a house, a car and have other items on hire purchase or time payment. This factor has caused the subdument of many an adventurous man. The load of hire purchase. With the inflated price of many items the "average everyday man in the street" cannot afford to save money for long periods of time to purchase it. Instead of that he gets credit and pays it off weekly.

This payment must be maintained or he loses the item. This factor will tie down many a person from changing his job or life style. Therefore he gets stuck in a financial rut, being unable to change lest he loses everything he has been working for. Also in many cases by the time he has finished paying for one item it requires replacement and so the vicious circle continues.

And so many Australians don't want to risk losing many of the items they have to change a job or try some new occupation or protest against unjust treatment.

When a politician who is earning 3 or 4 times the average weekly wage looks upon the majority who sit back and be silent and calls them soft he is judging a book by its cover, he is looking at one side of an argument. Many people cannot afford to protest or change as he can, the politician on many occasions does not look at people's financial status because he considers people to be similar to him. But they are not. His income is large and stable while other people must skimp and save to have a comfortable living. So, Mr. Politician look past your wallet and see that Australians are not "Soft" they are merely confined by credit companies, But nevertheless they are surely as "Hard" as you and if not Harder.

ESSAY X: STYLE I (ACTIVE/SAXON)

A member of parliament said lately that "Australians are Becoming Soft". This brings a thought to mind, Is the Australian image softening along with its people? The most common image of an Australian that people hold is a beer guzzling, quick tounged person. This image is surely a small section of the people and so this by itself should not be the basis for a wide-spread image.

Today's Australian lives basically in a suburb and spends most of his time withing an area arround his town or city. Thus this also turns upside-down the old Australian picture of "Man on the land".

So we have a picture of today's Australian image. A towny who lives and works within the city during the week and takes it easy at home on the week-end. He may be buying a house, a car with his pay and have other things on time payment. This point has quiet-downed many a daring man. The load of his time payment. With the blown-up price of many things the "everyday man in the street" cannot save long enough to buy it. Instead of that he goes into the red and pays it off weekly.

He keeps up payments or loses the goods. This will tie many a person down. He cannot change his job or way of living. So he keeps himself stuck in a rut with money, as he cannot change in case he loses everything he is working for. Also in many cases by the time he has ended paying for one thing he has to replace it and so the vicious circle goes on.

And so Many Australians don't want to risk losing many of the things they have to change a job or try some new work or speak out against unfair dealings.

When a member of parliament who is earning 3 or 4 times the average wage looks upon most people who sit back and be quiet and calls them soft he is judging a book by its cover, he is looking at one side of a case. Many people do not have the money or chance to speak out or change as he can, the member of parliament in many cases does not look at people's money position because he thinks people are like him. But they are not. He earns a big and steady income while other people skimp and save to be fairly well off. So, Mr. Member of Parliament look past your wallet and see that Austratians are not "Soft" money lenders just keep them in the one place, But nevertheless, they are surely as "Hard" as you and if not Harder.

ESSAY X: STYLE II (ACTIVE/LATINATE)

A politician remarked recently that "Australians are Becoming soft". This conjures up the notion, Is the Australian image softening in accordance with the population? The universal concept that people have formed of an Antipodean is a vociferous individual who consumes beer gluttonously. This image is undoubtedly a minority of the populace and therefore this should not constitute the sole foundation for a popular concept.

The modern Australian lives primarily in a suburb and expends the major portion of his time withing close proximity arround his town or city. Additionally therefore this phenomenon negates the long-perpetuated Australian concept of the "Rural dweller".

So we possess some conception of the modern Australian image. An urbanite who lives and labours within the metropolis throughout the week and relaxes at home on Saturdays and Sundays. With the financial rewards for his labours he may be purchasing a residence, a vehicle and other items that he commits to hire purchase. This factor has caused the subdument of many an adventurous head-of-family. The magnitude of hire purchase. With the inflated price of numerous commodities the "average citizen" cannot accumulate surplus financial resources over extensive periods of time to purchase it. As an alternative to that he seeks credit and reduces his principal in weekly instalments.

He must needs maintain his instalments or he surrenders the item. This factor will inhibit many an individual. He cannot vary his employment or modify his life style. Thus finance stagnates his existence, lest he sacrifice all that he has been attempting to achieve. Moreover in numerous instances by the time he has completely honoured his debt for one commodity he needs to purchase another and the vicious circle continues.

Consequently then Numerous Asutralians have no desire to hazard many of the items they possess to vary their occupation or attempt some new employment or protest against unjust treatment.

When a politician who receives 3 or 4 times the average wage surveys the majority of citizens who avoid publicity and remain silent and labels them as soft he is forming speculative conclusions on the slightest of evidence, he is conducting a unilateral argument. Substantial numbers of persons cannot afford to protest or vacillate as he is able to, the politician on numerous occasions does not examine people's financial status because he considers that people are similar to him. However they are not. He receives a large and stable salary while other people must manage their economic affairs most fastidiously in order to exist in some measure of comfort. So, Mr. Politician, extend your vision beyond your own coffers and acknowledge that Austratians are not "Soft" the credit companies merely confine them, But nevertheless they are undoubtedly as "Hard" as you and if not Harder.

ESSAY X: STYLE III (PASSIVE/SAXON)

It was said by a member of parliament lately that "Australians are Becoming Soft". A thought is brought to mind here, Is the Australian image softening along with its people? The most common image of an Australian is a beer guzzling, quick tounged person. A small section surely of the people is this image and so a wide-spread image should not be based on this alone.

Today's Australian is basically suburban and most of his time is spent withing an area arround his town or city. Thus the old picture of the Australian as a "Man of the land" is turned upside-down.

So a picture of today's Australian image is before us. A towny whose life and work during the week are within the city and whose week-end is at home taking it easy. A house, a car may be being bought with his pay and other things on time payment. Many a daring man has been quiet-downed by this point. The load of time payment. With the blown-up price of many things there is no way that saving long enought to buy it can be done by the "everyday man in the street". Instead of that going into the red and paying it off weekly is what he does.

This payment must be kept up or the goods are taken from him. Many a person is tied down by this from changing his job or his way of living. So he gets stuck in a rut with money, being unable to change in case everything that he has been working for is lost. Also in many cases by the time one item has been fully paid for it has to be replaced and so the vicious circle is kept going.

And so The risk of losing many of one's items to change a job and try some new work or speak out against unfair dealings is not wanted by many Australians.

When most people who sit back and be quiet are looked upon and called soft by a member of parliament whose earnings are 3 or 4 times the average wage this is the judging of a book by its cover, the looking at one side of a case. There are many people with no money and no chance to speak out or change as can be done by him, in many cases people's money position is not looked at by the member of parliament because it is his idea that there is a likeness between people and himself. But this is not true. His income is big and steady while skimping and saving must be done by other people to be fairly well off. So, Mr. Member of Parliament your wallet must be looked past and Austratians seen not as "Soft" they are just kept in the one place by money lenders, But nevertheless they are surely as "Hard" as you and if not Harder.

ESSAY X: STYLE IV (PASSIVE/LATINATE)

It was remarked recently by a politician that "Australians are Becoming Soft". A notion that is conjured up here, Is the Australian image softening in accordance with the population? The universal conception of an Antipodean is a beer gluttonous, voiciferous individual. Undoubtedly only a minority of the populace is this image and therefore a popular concept should not be founded solely on this.

The modern Australian is primarily suburban and the major portion of his time is expended withing close proximity arround his town or city. Therefore the long-perpetuated concept of the Australian as a "Rural dweller" is negated by this phenomenon as well.

So some conception of the modern Australian image is ours. An urbanite whose life and labours throughout the week are within the metropolis and whose Saturdays and Sabbath are for relaxation at home. From the financial rewards for his labours a residence, a vehicle may be in the process of being purchased and sundry commodities may be under hire purchase commitments. The subdument of many an adventurous head-of-family has been caused by this factor. The magnitude of hire purchase. With the inflated price of numerous items the accumulation of surplus financial resources over sufficiently extensive periods of time in order to purchase it cannot be achieved by the "average citizen". The alternative to this that is sought is the credit loan and reduction of principal by weekly instalments.

These instalments must be maintained lest the commodity be surrendered. Numerous individuals have by this factor been restrained from pursuing variety of employment or the modification of one's life style. Therefore there is a stagnation of existence financially, a constraint on innovation lest all that is being achieved is sacrificed. Moreover in numerous instances by the

time the debt for one commodity has been honoured in its entirety the item must be substituted by another and thus the vicious circle is continued.

Consequently then The hazard of forfeiting many of one's possessions to vary an occupation or attempt some new employment or protest against unjust treatment is antithetic to the ambitions of large numbers of Australians.

When the majority of citizens who avoid publicity and remain silent are surveyed critically and labelled soft by a politician whose salary is 3 or 4 times the average wage this is speculation and judgement on the slightest of evidence, the conduct of a unilateral argument by him. For large numbers of persons it cannot be afforded to raise protestations or to vacillate at will as is his prerogative, on numerous occasions people's financial status is not rationally acknowledged nor examined by the politican because it is his consideration that there survives a similarity between countrymen and himself. However the contrary exists in the extreme. His salary is large and stable while frugal and fastidious economies must be practised by other people to contrive an existence with some measure of comfort. So, Mr. Politician your vision must be extended beyond the compass of your own coffers and Austratians must be acknowledged as not being "Soft" they are merely confined by credit companies. But nevertheless they are undoubtedly as "Hard" as you and if not Harder.

ESSAY 2: ORIGINAL

Australians have always been known as the rugged masculine out door type person. But do they really live up to their image now days? Or has that day finished and now they are just like most European type persons, spoilt and lazy.

All you have to do is look around you, at this very moment, what do you see? a tough rugged man or just a plain ordinary soft person, with his comfortable office job or his $200 dollar suit, or maybe you might see a labourer working on the road, with his beer "gut" bulging out of his shirt.

Not many Australian men have the time or effort to keep up their image its either to hard or their girl friends wouldn't like it. Seeing that most Australians are getting the feeling of a comparitable easy life they don't feel like going back to their old style of life.

In Many ways the Australian person can not go back to his old style of life because of the social pressures that are placed on him.

Many Australians like to be known as tough and the out door type because it makes them feel good inside but it also makes the people feel guilty for leting down their country. Also many Australians don't want to be known as their image type because most foreigners may act as though they think a lot of them but really they are laughing inside them at their stupidness.

There may of been a tough masculine out door type Australian about 10 years ago but now days there are only a few left and most of these are stupid and ignorent.

I must say from just looking around that it is pretty hard to find a typical Australian that is masculine rough out door type because it is easier to be different.

ESSAY Z: STYLE I (ACTIVE/SAXON)

We have always known Australians as the rugged manly out door type person. But do they really live up to their image, now days? Or has that day gone and now they are just like most European sort person, spoilt and lazy.

All you have to do is look around you, right now, what do you see? a tough rugged man or just a plain everyday soft person, with his easy office job or his $200 dollar suit, or maybe you might see a man working on the road, with his beer "gut" bulging out of his shirt.

Not many Australian men have the time or effort to keep up the way they should be they find it to hard to do or their girl friends wouldn't like it. Seeing that most Australians are getting the feeling of living pretty-like easy they don't feel like going back to their old way of life.

In Many ways the Australian does not go back to his old sort of life because people around him wouldn't.

Many Australians like to be known as tough and the out door type because that makes them feel good inside but they also feel guilty for leting down their country. Also many Australians don't want people to see them as being like the image because most people from overseas may seem freindly and act as though they think a lot of them but really they are laughing inside them at their stupidness.

The Australian may of been a tough manly out door type about 10 years ago but now days we see only a few of them and most of these know nothing and have little sence.

I must say from just looking around that I think it is pretty hard to find a typical Australian that is manly rough out door type because we find it easier to differ.

ESSAY Z: STYLE II (ACTIVE/LATINATE)

One has continuously in the past acknowledged Australians as the rugged masculine type person who experiences life al fresco. But do they in point of fact support and honour this conception of them, at this pointed time? Or has that person declined and presently are they to all appearances similar to the majority of European model person, pampered and slothful.

You have no greater obligation than simply to glance cursorily about your vicinity, at this very moment, what do you observe? a tenacious, rugged individual or simply an unimpressive ordinary soft person, with his comfortable sedentary occupation or his $200 dollar suit, or perchance you have the opportunity of observing a labourer toiling on the motorway, with his alcoholic corpulence billowing out of his shirt.

Relatively few Australian males find time at their disposal or apply themselves vigorously to the pursuit of maintaining the popular conception of themselves they discover either that its to difficult or that their sweethearts would not approve. In view of the fact that the majority of Australians are becoming increasingly acquainted with living a comparitable comfortable life they do not relish the notion of regressing to their former style of life.

By and Large the Australian does not seek the earlier style of life as a result of the fact that social pressures constrain him.

Numerous Australians receive with pleasure the acknowledgement of being tenacious and the naturalistic type because that sparks a glow of extreme satisfaction within but in addition they experience self-culpability for discreditting their nation. In addition numerous Australians prefer not to be recognised as reflecting the image because the majority of foriegners may behave as though they respect them but actually they are sneering within them at their stupidness.

The Australians may of been a tenacious masculine al fresco type about a decade ago but at this pointed time we can account for but a limited number remaining and the majority of these are stupid and ignorant.

I must conclude as a result of simply surveying the scene that I consider it rather difficult to discover a typical Australian that is masculine unrestrained naturalistic type because we can vary with less discomfort.

ESSAY 2: STYLE III (PASSIVE/SAXON)

Australians have always been known as the rugged manly out door sort person. But is their image really lived up to, now days? Or has there been an end to that day and now are they just like most European type person, spoilt and lazy.

All that has to be done is look around, right now, what is seen? a tough rugged man or just a plain everyday soft person, withhis easy office job or his $200 dollar suit, or maybe a man might be seen at work on the road, with his shirt being bulged out by his beer "gut".

No time or effort is there for their image to be kept up by many Australian men its to hard or it wouldn't be liked by their girl friends. Seeing that getting the feeling of living pretty-like easy is done by most Australians there is no feeling like going back to their old way of life.

In Many ways the Australian person is not able to go back to his old way of life because of what would be thought by the people around him.

There is a liking by many Australians of being known as tough and the out door sort because they are made to feel good inside but they are also made to feel guilty for leting down their country. Also to be known as the image type is not wanted by many Australians because they look as though they are freindly with and thought a lot of by most people from overseas but really inside them is laughter at their stupidness.

There may of been a tough manly out door type Australian about 10 years ago but now days there are only a few left and most of these are senceless and know very little.

It must be said from just a look around that it is pretty hard to find a typical Australian that is manly rough out door type because it is easier to be different.

ESSAY Z: STYLE IV (PASSIVE/LATINATE)

Australians have continuously in the past been acknowledged as the rugged masculine type person whose experience is of living al fresco. But in actual fact is this conception of them supported and honoured by them, at this pointed time? Or has there been the demise of that era and at present are they to all appearances similar to the majority of European model person, pampered and slothful.

It is necessary only to glance cursorily about your immediate vicinity, at this very moment, what is to be observed? a tenacious, rugged individual or simply an unimpressive ordinary soft person, with his comfortable sedentary occupation or his $200 dollar suit, or perchance a labourer might be perceived toiling on the motorway, with his shirt billowed by his alcoholic corpulence.

For relatively few Australian males is there time sufficient to be utilised or energy to be expended on the maintenance of the popular conception of them its to difficult or it would fail to be approved by their sweethearts. In consideration of the fact that increasing acquaintance and familiarity with a comparitable luxurious life has been experienced by the majority of Australians the notion of regression to their former life style is not relished.

By and Large the Australian citizen is in the situation of being prevented from seeking refuge in the earlier style of life as a consequence of the social pressures on him.

There is pleasure for numerous Australians to be acknowledged as being tenacious and the naturalistic type because of a glow of extreme satisfaction is sparked within them but in addition self-culpability is experienced by them for discreditting their nation. In addition to be acknowledged as the epitome of the Australian image is not the desire of numerous Australians because feigned respect is bestowed upon them by the majority of foriegners when in actuality there is a sneer within them at their stupidness.

There may of been a tenacious masculine al fresco type Australian about a decade ago but at this pointed time there are but limited number of them remaining and the majority of these are stupid and ignorent.

It must be concluded from a simple survey that it is rather difficult to discover a typical Australian that is masculine unrestrained naturalistic type as a concequence of the reality that there is less discomfort in variation.

APPENDIX B

INSTRUCTIONS TO GRADERS TOGETHER WITH GRADES AND COMMENTS MADE BY EXAMINERS IN PILOT STUDY

INSTRUCTIONS TO ESSAY-GRADERS

Grade the following essays on a 5-point scale. Use the scale A , B , C , C- , D .

The equivalent marks out of ten for each letter grade would be approximately like this, if you prefer to think numerically:

D	C-	C	B	A
Bad Fail	Just Fail	Satisfactory	Good	V. Good/Excellent
0 - 3½	4 - 4½	5 - 6	6½ - 7½	8 - 10

PLEASE FOLLOW THIS PROCEDURE:

1. Use rapid general impression to award a letter grade for the essay over-all, and record the grade on the Cover Sheet for that particular essay.

2. After arriving at the general, over-all grade, reconsider the essay's Mechanics, Usage and Grammar. Award a grade just for this.

3. Next, award another grade for the Organisation and Logic of the essay.

4. Write a short comment about the essay's strong or weak points, or your reactions to it.

The essays are in a bundle in random order for marking.

MARKER: 1

MARKS SHEET

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
V.I	B	B	B	The writer has expressed himself clearly and accurately. Paragraph development is rather weak, but generally, the essay drives accurately and smoothly forward
V.II	A	A	A	The student shows an excellent command of the language. The principal weakness though not particularly serious here, is a tendency towards repetition.
V.III	C	C	C-	This student is very erratic. He swings from good common sense to nonsense. However, there is enough of value in the essay to be worth a pass mark.
V.IV	A	B	B	Again, a fair proportion of the essay is second-hand, but he has made maximum use of the material.
W.I	C-	C-	C-	The grammar and logic are too weak to be worth a pass. There seems to be hardly two consecutive sentences which are fully correct. There are far too many generalizations.
W.II	B	C	B	The essay's logic varies from sound reasoning to vague non-sequiturs. Good use has been made of vocabulary. The grammatical blunders seemed to be made more from haste than ignorance.
W.III	C	B	B	While the essay is fairly sound in organisation and grammar, nothing very new emerges. The clichés of behaviour are too readily accepted by this student without any qualification at all.
W.IV	A	B	B	The student has written fluently and coherently, and apart from minor errors of haste, has written accurately. Rather much seems to have been taken directly from the essay topic, but he has made good use of the material.
X.I	C	C	C	The writer begins fairly well but swings away from what appears to be support for the Australian. There is little evidence of originality in this paper, words and phrases being taken directly from secondary sources.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
X.II	B	B	B	The essay reads quite well, but lacks originality. Paragraph structure tends to be rather loose, but generally makes for a reasonable progression.
X.III	C-	C-	C-	There is too much looseness of thought, too much acceptance of the clichés of Australian behaviour. Little real evidence has been provided to support the student's main statements.
X.IV	A	B	B	The student has approached the topic with a good deal of confidence, making as much use of secondary information as possible. The handling of periodic sentence structure is generally good.
Y.I	D	D	D	The student seems unsure of what he is trying to say. No theme has emerged from this tangle of doubtful generalisations.
Y.II	D	C-	C-	The student seems to be using words and phrases from the topic without knowing what they mean. The essay tends to be an answer to a question without sufficiently developing ideas.
Y.III	D	C	C-	The question does not seem to be understood fully. The line of argument is extremely tenuous. The writer would have been better off using his own common-sense rather than badly apply material of others.
Y.IV	C-	D	C	There is far too much word-spinning for the sound rather than the sense. The level of vocabulary used is good, but the manner in which it is used fails to completely hold together.
Z.I	D	D	C-	The student has included too many generalizations, which are extremely doubtful. There is too little evidence of clear development of an argument.
Z.II	C	C-	C	The essay sounds better than it is. The vocabulary is of a good level, but the student has applied the words too sweepingly with too little qualification.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
Z.III	D	C-	D	The essay lacks cohesion within the paragraphs and from one paragraph to the next. Hardly any evidence is offered in support of generalisations.
Z.IV	B	B	C	The student has strung his material together quite well, but it lacks originality. I begin to doubt whether he understands what he is saying.

MARKER: 2

MARKS SHEET

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
V.I	C	C	B	An overabundance of hyphenated words probably due to an inadequate vocabulary. Also overuse, and misuse of commas. Organisation is satisfactory.
V.II	A	B	A	Mature vocabulary, well organised and structured. Mechanics generally good, misuse of comma evident; spelling sound.
V.III	C	C	C	Tendency towards overuse of hyphenated words and phrases shows a lacking in vocabulary, eg faster-than-sound instead of supersonic. Sentences generally tend to be too long, leading to clumsy construction. At least twice, adverbs have been misused as adjectives, normally not very significant but here very noticeable in context. Opening and concluding paragraphs not satisfactory.
V.IV	A	A	A	Well organised essay. The minor grammatical mistakes are not of sufficient magnitude to detract from the overall standard.
W.I	B	C	B	Whilst organisation is satisfactory, the generally poor mechanics and slang detract from it. However content is good and this balances out the overall standard.
W.II	B	C	B	A good vocabulary indicated. Only a minor portion of our reasoning is done by computers. Hasn't really come to terms with different types of pressures, not only physical, which people are subjected to nowadays.
W.III	B	B	A	Vocabulary not very satisfactory eg "done" for committed. Slang is overused but the conviction of the writer is generally apparent as being genuine. Most arguments are well illustrated, although Alan Moorehead will be pleased to know of his "Fatal Impact". Generally any lack in mechanics is overcome by writer's sincerity.
W.IV	A	B	A	A lucid style; too lucid in parts where the thought processes haven't been established in order. A good vocabulary which is well used.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
X.I	C	C	C	Vocabulary is inadequate and content narrow. There is a hint of vitality but it is not sustained long enough to overcome the restricted framework within which the writer has chosen to write.
X.II	C	B	B	The credit companies can't take all the blame if Australians are becoming soft, perhaps they are born of a demand, not vice-versa. Content is narrow, but a good vocabulary, perhaps 'overdone' is evident.
X.III	C-	C-	C	Style is disjointed and not helped by too many hyphenated words and phrases, indicative of an inadequate vocabulary. Syntax and sentence structure are weak. Once again the money-lenders can't take all the blame.
X.IV	B	B	B	Sound vocabulary, tendency to overuse the "big" words. Sentences often incomplete and not used correctly.
Y.I	D	D	D	Little coherence if any. The long, last paragraph is irrelevant as is the maxim with which the writer ends.
Y.II	C-	C	C-	An adequate vocabulary is displayed despite the inaccurate spelling of common words. Again the last half of the essay seems irrelevant.
Y.III	D	C-	C	Logic and organisation are fair in so much as the writer stays with the one restrictive and erroneous aspect of the topic. The whole democratic world can hardly be blamed if Australians are becoming soft.
Y.IV	C-	C	B	A significant proportion of the essay is irrelevant, brought about by the very narrow outlook taken by the writer. Logic and organisation are reasonable in so far as the writer has continued with his narrow argument.

Essay	Over-all Grade	Usage/Grammar	Organisation/ Logic	Comments
Z.I	C-	C	C-	Repetitive and superficial. Vocabulary is limited. Not enough thought is given to the topic, or else writer has chosen a narrow restricted framework.
Z.II	C	C-	C	Mechanics, especially usage and grammar are not good. Same vivid writing, e.g. "alcoholic corpulence billowing", compensates partly for substandard usage, but not enough to warrant a pass mark.
Z.III	C-	C-	C-	Repetitive, limited vocabulary, little unity or coherency.
Z.IV	C	C	B	Perhaps too restricted in content, mechanics are only just satisfactory, while organisation is fair.

MARKER: 3

MARKS SHEET

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
V.I	B	C	B	Some inaccuracies, use of the wrong word, or inadequate expression, interferes with a generally well-structured essay. Some more specific points to support generalizations would also help.
V.II	B	B	A	While the ideas from this writer come through quite clearly some of his phrasing contains a lack of harmonised development, which interferes with a complete appreciation of his point of view.
V.III	C-	C-	C	Confusion results from inaccuracies of language and an excessive generalisation.
V.IV	A	B	A	An obviously keen and agile mind has assembled a very useful range of material. Some of it is not given sufficient space for full development and therefore parts of the essay lack clarity as the ideas tumble over each other.
W.I	C	C	B	Standard of expression improved as the essay developed and the writer gained a better sense of perspective for his argument.
W.II	C	C	C	Clarity is affected by long paragraphs including heavy and doubtful generalizations. Thematic unity is not easily held and the conclusion is inadequate.
W.III	B	B	B	The many ideas, which gush out, are in fact generally worthwhile. Better paragraphing and structure of the essay would make it one of the best. At least the communication is sound because of mechanical expertise.
W.IV	A	B	A	High command language and conceptual awareness. Some flaws in presentation are of minimal importance in the student's grasp of the subject.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
X.I	C	C	B	Colloquial expressions have some attraction, but lower standard of communication for this type of essay.
X.II	B	C	B	The writer has a sense of occasion in the way that he builds up a pressure essay, reflecting the average man he is describing. The writing also contains conviction and maturity.
X.III	C-	C-	C	Although the logic of the argument is sound, confusion is created by the inaccuracies of the writer's expression.
X.IV	C	C	B	Better paragraphing and the use of more simple language would have improved the communication level of this essay considerably. The writer seems to be trying to write impressively rather than clearly.
Y.I	C-	C-	C	A satisfactory logic in presentation is not sustained in the quality of other aspects of the writing.
Y.II	D	D	C-	Inaccuracies of expression contribute greatly to lack of thematic unity. Ideas flow too fast for his organisational ability.
Y.III	C-	D	C-	Further development, based on more detailed organisation, could result in a sound essay.
Y.IV	D	D	C-	Some logical thread of unity is evident, but inadequate mechanics cause the general standard to be lowered.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
E.I	D	D	D	Possibly logical sequences are destroyed by inaccuracies and unfinished statements. Gross confusion results.
E.II	C-	C-	C-	Although isolated parts of the essay are effective the overall effect is one of confusion and lack of precision about what the writer really wants to say.
E.III	D	D	D	Sweeping generalizations in confused sentences leave the reader with a very scrambled impression of the writer's ideas.
E.IV	C-	C-	C	There seems to be a quality of mind and appreciation of language, struggling for recognition, but hindered by lack of control of the mechanics of writing. Some terms of phrase are quite delightful.

175

MARKER: 4

MARKS SHEET

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
V.I	B	A	A	Well developed theme of how emotional strength has had to replace physical strength. Fluent, but a tendency to list-like repetitions. No real examples given of emotional stress, however. Vocabulary is reasonably mature.
V.II	A	A	B	A well-developed theme but a questionable conclusion. Emotional hardening sounds frightening. Sound vocabulary and fluent expression.
V.III	A	B	A	Concept of emotional strength well developed but "keeping feelings out" is not a sensible conclusion. Fluent and forceful expression, adequate vocabulary.
V.IV	B	C	B	Syntax is clumsy at times but exhibits a mature vocabulary relevant to the topic. Impressive logical development except for the linking of "survival technique" to softening which is not made clear.
W.I	B	C	A	Impressively developed, each paragraph supporting the main theme by example and explanation. Only a few solecisms.
W.II	B	B	A	Main feature is the educated but overdone and hence pretentious vocabulary. Some circumlocutory sentences but paragraph development is sound as are development of ideas and theme.
W.III	A	B	A	Impressively developed theme. Language and expression is "down-to-earth" though formal. Good use of supporting detail.
W.IV	B	B	A	Well developed ideas. Fluent but pretentious vocabulary.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
X.I	C-	C	D	Impressive sincerity, colloquial expression with a few too many clichés. Logical development lacking on financial commitments, tied to "softness".
X.II	C	B	D	Sound vocabulary would indicate a reasonably capable student which does not correlate with the lack of logic. Much of the content on hire-purchase does not develop the "soft" Australian theme.
X.III	C-	C	D	Logic is at fault. The financial problems have not been clearly related to the "soft" concept. The non-sequitur again, it doesn't follow that a man with his financial burdens is soft. Uses colloquial language and clichés although reasonably fluent.
X.IV	C	B	C-	Inversions and emphasis on pedantic vocabulary produces very stilted expression. Vocabulary indicates a maturity but the linking of "Soft Australians" and their financial resources is not convincing logically.
Y.I	C-	D	C	Ideas are reasonably developed from his opening concept. Mechanics (spelling especially) are weak but sentences are reasonably fluent.
Y.II	B	C-	B	Topic is well explained and limited. Maturity of vocabulary is difficult to associate with poor spelling and non-sentences. His theme is strong and communicates effectively.

Essay	Over-all Grade	Mechanics Usage/Grammar	Organsiation/ Logic	Comments
Y.III	D	D	D	Does not develop a clear idea of what becoming "soft" means hence his example becomes a non-sequitur. He shows obvious concern for small businesses, but not made this relevant.
Y.IV	C	C-	C-	A maturity of vocabulary but poor spelling. Opening paragraph is impressive as is topic sentence of last paragraph. Final example is a valiant attempt to refute Mr Killen. Hence is a logical response to the accusation for the writer who obviously understands business principles.
Z.I	C	C	C-	Simplified view but consistently developed. Colloquial and conversational ideas of paragraphing are poor.
Z.II	C-	C-	D	Sentences contain unrelated ideas are vague and circumlocutory. An artificial style that doesn't communicate with clichés "this pointed time", "by and large".
Z.III	C	D	C-	Some awkward expression, syntax is clumsy - ideas confused, lack of logical development, ideas are limited.
Z.IV	C	C	C-	Many sentences either too vague to be meaningful or have words misused. Rambling sentences with a pretentious use of vocabulary haven't helped clarity or logical development. Some interesting concepts and images.

MARKER: 5

MARKS SHEET

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
V.I	B	B	B	A reasonably clear argument, lucidly written, although there are contradictory elements.
V.II	A	A	B	Repetitive in parts and rather contradictory, but reasonably well written.
V.III	C	C	C	--
V.IV	B	B	B	--
W.I	C	B	C	Some very obvious contradictions.
W.II	B	B	B	An irritating tendency to verbosity, generally thoughtful but tends to repetition.
W.III	B	C	B	Points reasonably well made, sentence structure weak at times.
W.IV	B	B	B	Less wordy and more coherent than some previously on this pattern, but there is still an unnecessary verbosity.
X.I	C	C	C	A poorly structured argument, none too well expressed.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
X.II	B	B	B	The argument is more clearly expressed than in previous "models". The phraseology still rambles, though it is far more moderately used by this writer.
X.III	C	C	C-	A woolly argument with not much relevance to the topic.
X.IV	C	C	C-	Again an imitating use of "verbage" - neither is there much logic in the argument.
Y.I	C-	C	C-	Too similar to essay 16. Collaboration?
Y.II	C-	C	D	Uncertainness is certainly generating and spralling. Again I cannot follow the "logic" of the argument.
Y.III	C-	C	C-	My previous comments on this essay apply.
Y.IV	C-	C	C-	An odd interpretation of the topic. Seems to operate on the theory that if you use enough polysyllabic words, the sense, or lack of it, will be obscured.
Z.I	C-	C-	C-	The ideas are there but they are poorly expressed.
Z.II	C	C	C	My reaction is that this student has not made very good use of the material presented to him. Words and phrases appear to have been strung together at times with little appreciation of meaning.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
Z.III	C	C-	C	Poor expression
Z.IV	C	C	C	The expression in this essay and others of the same pattern is that of a lesson indifferently learned and poorly regurgitated. Very difficult to assess.

MARKER: 6

MARKS SHEET

Essay	Over-all Grace	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
V.I	B	B	B	Contrast between physical deadline and emotional toughness could be more fully developed to give greater cohesion to the essay.
V.II	C	B	C-	Roundabout style of argument. Weak on linking from one idea to another. Little sense of personality of writer.
V.III	C	C	C	Peculiarities of syntax. Uncertain use of pronouns. Peculiar logic that emotional strength is more significant in process of staying alive than is physical strength - point is not made clearly.
V.IV	B	B	C	Peculiarity of logic - equating comfort with survival.
W.I	B	B	C	Logic of higher standard of living causing more people to go on the dole, or cheat social service institutions is a bit confusing.
W.II	A	B	A	Tendency to try to cram too much into one sentence. Some points of argument have become overburdened by examples.
W.III	C	C-	C	Peculiarities of syntax and sentence structure. Some attempt to link cause and effect but some of the links eg between higher standard of living and dole-cheats, seem fairly tenuous.
W.IV	A	B	A	Occasional lapses in mechanical accuracy which seem odd when vocabulary, sentence development, logic is taken into account.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
X.I	C	C-	C	Idea that the image of the Australian has changed could be developed more fully. The contrast of the Parliamentarian's situation with that of the ordinary man's is not clearly linked with the opening idea.
X.II	B	B	B	Concept of a change in the image of the average Australian is not clearly linked with the contrast between the condition of the average Australian and the politician. Handling of the final sentence is odd - the opening rhetoric suggests a competence with language that is not supported by the close of the sentence.
X.III	C	C-	C	Changing image of the Australian is not clearly related to the second part of the argument. Sentence structure is shaky.
X.IV	A	A	A	Complexity of language tends to obscure the point being made, in many instances.
Y.I	C	C-	C	A certain logic in the development of argument is not clearly related to the concept of "softness". Spelling, sentence structure and sentence development are poor.
Y.II	C	D	B	The logic displayed in the course of the argument is not connected firmly enough with the opening paragraph. Peculiarities of spelling and sentence structure do not seem consistent with the vocabulary level.
Y.III	C	C-	C	Relationship between the idea that the Australians are becoming too soft and the economic condition of the country is not drawn very clearly.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
Y.IV	C	C-	C	Relationship between the economic situation and the dependence of people on Government subsidies is not made very clearly.
Z.I	C	C-	C-	Sentence structure and use of rhetorical questions shaky. Generalizations about the European way of life and the intelligence of the outdoor type of Australian unsupported by evidence.
Z.II	B	C-	C	Remarkable vocabulary - tends to be elevated in tone - probably intended to be satirical but can't sustain it. Lively style of writing, but hasn't mastered the mechanics to go with it.
Z.III	C-	C-	C-	Sentence structure poor. Generalisations "just like most European-type persons" and "most of these are senseless" are unsupported by any evidence, and have little to do with the argument.
Z.IV	C	C	C	Vocabulary is not supported by facility with mechanics of expression. Generalisations about the quality of European life, or the intelligence of the "al fresco" Australian unsupported by facts.

MARKER: 7

MARKS SHEET

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
V.I	A	A	A	Pretty good all round - the argument is simplified and developed. It's a little repetitive towards the end, and there are a few clumsy expressions or stylistic carelessness, but it's a good effort (A- or B+).
V.II	A	A	A	Pretty good all round - probably the best of the bunch.
V.III	B	B	C	A little repetitive, and lacking in cohesive development, especially the last half, otherwise not bad.
V.IV	A	B	A	I'd rather give a B+... A good argument - but a little repetitive in the last paragraphs. Good usage - but verging on verbosity. Good style, but a little abstract and pretentious.
W.I	B	B	A	Well organised and developed case - usage and mechanics are very variable, with occasional significant weaknesses, but the essence comes through strongly.
W.II	B	B	A	Uneven effort - in parts, usage is very good; in parts, it is clumsy, awkward and pretentious. Organisation is generally very good, however, even if there are repetitive elements.
W.III	B	B	A	Rather give B+ or A- ... Pretty sound in all respects, save for some awkward and clumsy (careless?) expression and usage in places.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
W.IV	A	B	A	Intelligent and well-developed essay. Errs on the side of pretentiousness to some degree - actually says less than it seems to say, and is socially clichéd somewhat.
X.I	B	C	B	Usage is very average, but the writer makes a reasonable case - although that case diverges from my reading of the topic. C to B is a better mark.
X.II	B	B	A	The essay's second half is better handled than the first - the latter containing many more glaring errors, as well as a greater degree of grandiloquence of language. The argument is well presented, if to use his own borrowed word 'unilateral'.
X.III	B	C	A	Usage C+ or B-? - very variable and much better early than late in the essay. It at least develops a good argument for the most part - some clumsiness towards the end, a concrete, flowing style.
X.IV	C	C-	B	Beautifully awkward, formal and (impersonal (passive style! Good argument, nearly lost in the verbiage.
Y.I	C-	D	C	As in other essays of this ilk, the argument is at least there to be seen, but still seems to be largely to be irrelevant. Usage is poor, clumsy.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
Y.II	B	C	B	Struggling hard with a complex idea, and builds well from his thesis, but it's a bit much for him and he cannot handle his medium quite as well as his idea.
Y.III	C-	D	C	Given that the student has diverged from his stated topic, he at least pursues his deviation with some logic. Otherwise, not much good.
Y.IV	C-	D	B	Deviation from the point - as I see it - but some sort of case is at least developed. Usage, style - poor. Org./logic really worth C+ or B-.
Z.I	C-	C-	C-	Very ordinary - not quite as contradictory as some essays in the chosen expression of dubious argument, but not much good in any aspect.
Z.II	C	C	B	An element of pomposity and pretention. An argument is developed, and is at least recognisable through the less-than-well-handled verbosity.
Z.III	D	D	D	Weak - contradictory in argument, poor in expression.
Z.IV	D	D	C	Eh? If there is a real argument, it is too disguised beneath grandiloquent, pompous and meaningless verbosity - poorly handled at that. The writer doesn't communicate.

MARKER: 8

M A R K S S H E E T

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organsiation/ Logic	Comments
V.I	B	B	B	This appears very similar in content to an essay I marked earlier, but not so well expressed.
V.II	A	A	A	The writer has effectively communicated clearly his point of view.
V.III	B	B	A	The writer has communicated clearly his moderate point of view.
V.IV	A	A	A	The writer has argued clearly and from an interesting point of view. There are few errors in mechanics.
W.I	C	C-	C	The writer is very much involved with the topic, communicating effectively despite some errors in mechanics.
W.II	A	B	A	The writer has a good understanding of the topic and writes fluently.
W.III	B	C	B	This just makes "B" grade. The topic is well covered and the expression is satisfactory.
W.IV	A	A	A	Articulate and mature.
X.I	C-	C-	C-	The writing lacks cohesion. There are numerous errors in mechanics.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
X.II	C-	C-	C-	This was a rather "strange" essay. Not only did the writer stray from the topic but there is considerable misuse of language. Perhaps the verbose style results from a recent reading or discussion in class on topic related to the essay one. The writer is not master of the expression.
X.III	C-	D	D	This is a borderline D/C-. There is a lack of originality and the expression is weak.
X.IV	C-	C-	C-	The expression is stilted and there are numerous errors in mechanics. The language appears unnatural for the student. This essay is barely C-.
Y.I	D	D	D	The writer lacks personal opinion and is repeating something that he has been told. There are numerous errors in mechanics.
Y.II	D	D	D	There are numerous errors in his mechanics as well as failing to write on the topic.
Y.III	D	D	D	The writer has been distracted by the economic situation and has failed to answer the question. There are numerous errors in mechanics.
Y.IV	D	D	D	This essay is almost identical to Essay No. 15. The same comments apply.

Essay	Over-all Grade	Mechanics/ Usage/Grammar	Organisation/ Logic	Comments
Z.I	C	C	B	Too brief to be considered "B" Grade.
Z.II	C	C	C	There are no noticeably strong points. It doesn't warrant the "just Fail" Grade.
Z.III	C-	C-	C	This essay lacks the clarity of expression and sufficient content of argument for it to be classified higher.
Z.IV	A	B	A	The writer is articulate and shows keen insight.

APPENDIX C

THE USE OF THE

RASCH POLYCHOTOMOUS RATING MODEL

APPENDIX C

THE USE OF THE RASCH RATING MODEL*

The extended model for ordered responses takes the form

$$p\{x_{vi} | \beta_v, \delta_i\} = \frac{e^{\kappa_x + x_{vi}(\beta_v - \delta_i)}}{\sum_{k=0}^{m} e^{\kappa_k + k(\beta_v - \delta_i)}}$$

where x_{vi} is the rating response which can take the values $0,1,2,\ldots,m$, where in the present case $m=4$, β_v is the parameter associated with essay v and can be taken to characterise its "ability" or quality, δ_i is the parameter associated with grader i and can be taken to characterise the grader's harshness, and κ_x are category coefficients which are related to cut-off points or thresholds τ_k on a continuum according to

$$\kappa_o = 0$$

and

$$\kappa_x = -\sum_{k=1}^{x} \tau_k \,, \quad k=1, m.$$

Parameter Estimates

The equations which must be solved for the parameter estimates are

$$r_v = \sum_{ix} \sum e^{\kappa_x + x(\beta_v - \delta_i)/\gamma_{vi}} \,,$$

$$s_i = \sum_{vx} \sum e^{\kappa_x + x(\beta_v - \delta_i)/\gamma_{vi}} \,,$$

and

$$T_x = \sum_{iv} \sum e^{\kappa_x + x(\beta_v - \delta_i)/\gamma_{vi}} \,,$$

---

* This Appendix was written with the help of David Andrich.

where $\gamma_{vi} = \sum_{k=0}^{m} e^{\kappa_k + k(\beta_v - \delta_i)}$ is the normalising factor,

$r_v = \sum_i x_{vi}$ is the total score of essay v,

$s_i = \sum_v x_{vi}$ is the total score of grader i,

and T_x is the total number of times that score x was given over all essays by all graders.

It should be noted that parameterising the graders in terms of harshness implies that it is not required that graders give the same ratings for reliability of grading to be evident. For evidence of reliability, it is sufficient that any differences among graders can be explained by differences in harshness and some random error given by the probability equation. That this is the case is checked by testing the fit of the data to the model.

Test of Fit

The test of fit involves using the parameter estimates in the model, predicting the ratings that would be expected if the model was appropriate, and then comparing the observed ratings with the predicted ones. How the model matches the observed data may be examined using several methods, but in this study, the technique in which graders are divided into class intervals was used. In particular, because of their relatively small number, graders were divided into two class intervals. The graders with total scores above the median score, which represented the lenient graders, were placed in one class interval, and the graders with total scores below the median were placed in the other class interval.

For each grader i_g in each class interval g, $g=1,2$, the expected rating, $E[x_{vi_g}]$, to each essay v can be obtained from

$$E[x_{vi_g}] = \sum_{x=0}^{4} \hat{p}_{xvi_g}$$

where $p_{xvi_g} = e^{\hat{\kappa}_x + x(\hat{\beta}_v - \delta_i)/\gamma}$.

Over all the graders in the class interval g, the total of ratings for any essay is given by

$\sum_{i=1}^{I_g} E[x_{vi_g}]$, and this is compared to the observed total of the ratings $\sum_{i=1}^{I_g} x_{vi_g}$, where I_g is the total number of graders in class interval g.

If the ratings conform to the model, then for each essay, the observed and expected values should be similar. The degree of similarity is evaluated by a chi-square test: if the chi-square test shows significance, then the ratings of the different graders to the essay are not conformable. If it does not show significance, then inter-rater reliability is indicated.

As reported in chapter V, the chi-square tests revealed that the model held in all applications. (See also Tables I - III in this Appendix.)

Ability Measures and Advantages of the Rasch Model Transformations

Ability measures for each essay v are represented by the parameter estimate $\hat{\beta}_v$. This is in logits which typically take values from -3 to 3. Therefore for convenience, these values were multiplied by 4.55, then had 50 added to them, and finally rounded to whole numbers. It should be pointed out that the score of 50 does not correspond to a threshold value between pass and fail.

The use of transformed scores of essay ability has an important advantage over the use of simple grade-scores as recorded by the graders initially. It is known that traditional measures of ability on tests are not really interval measures but, strictly speaking, ordinal measures (Kerlinger, 1976). The intervals between adjacent test scores in the middle of a distribution do not represent the same increments in ability as the intervals between adjacent scores at the tails of the distribution. In the case of essay gradings, the difference in ability represented by the grades of 3 and 4 is not the same as the difference in ability represented by the grades 4 and 5. In the transformation from raw scores to transformed scores using latent trait models like the Rasch models, the transformations onto a logarithmic metric allow for additive properties to be incorporated into the parameters. The result is that differences in ability in the middle and at the extremes of the response distribution are directly comparable by virtue of the widening of the intervals at the extremes. The consequent effect on variances allows for more precise testing of experimental hypotheses.

These ability measures for each essay were then analysed according to the principles of multivariate analysis of variance design, recognising repeated measures structure of the data, to test the hypotheses of the study. The repeated measures feature arose because each original essay was rewritten into four different styles, and the original general quality of the essay would have carried through all its versions.

TABLE C I

CHI-SQUARE VALUES FOR OBSERVED AND EXPECTED PROPORTIONS OF 1977 GRADER AWARDS TO EACH GRADE-SCORE FOR ESSAYS ON OVERALL GRADING DIMENSION

Lenient Graders'	Mean Logit	.016
Medium Graders'	Mean Logit	.625
Harsh Graders'	Mean Logit	.953

**

ESSAY CODE	ABILITY LOGIT ESTIMATE	χ^2 (d.f. 1.92)	p<
V AS O	- .67	1.466	.46118
V AL O	-1.306	.983	.59169
V PS O	- .487	2.982	.21236
V PL O	-1.101	2.353	.29262
W AS O	.116	5.464	.06032
W AL O	-1.505	3.330	.17789
W PS O	- .190	2.968	.21392
W PL O	-1.616	2.602	.25774
X AS O	.003	1.562	.43886
X AL O	- .735	.787	.65535
X PS O	.049	1.355	.48831
X PL O	- .702	.866	.62895
Y AS O	1.309	1.923	.36461
Y AL O	.901	6.285	.03985*
Y PS O	1.241	2.889	.22270
Y PL O	1.089	2.360	.29172
Z AS O	.849	1.145	.54417
Z AL O	- .293	4.442	.10120
Z PS O	.761	.452	.78124
Z PL O	- .402	1.164	.53883
P 21 O	.160	1.496	.45401
P 22 O	.367	1.815	.38552
P 23 O	.634	1.611	.42792
P 24 O	.560	1.648	.41990
P 25 O	.970	.868	.62824
Joint Test of Fit (d.f. 48)		54.818	.23185

** For the 1977 analysis of fit it was appropriate to use three class intervals of grader "harshness" measures, not two.

TABLE C II

CHI-SQUARE VALUES FOR OBSERVED AND EXPECTED PROPORTIONS OF 1977 GRADER AWARDS TO EACH GRADE-SCORE FOR ESSAYS ON MECHANICS GRADING DIMENSION

Lenient Graders' Mean Logit	-.115
Medium Graders' Mean Logit	.453
Harsh Graders' Mean Logit	.945

ESSAY CODE	ABILITY LOGIT ESTIMATE	χ^2 (d.f. 1.92)	p <
V AS M	- .547	1.231	.52062
V AL M	-1.179	2.412	.28403
V PS M	- .264	2.779	.23549
V PL M	-1.435	3.620	.15356
W AS M	.268	.644	.70619
W AL M	-1.139	7.555	.02099*
W PS M	- .015	.346	.82667
W PL M	-1.139	4.117	.11931
X AS M	.229	.182	.90279
X AL M	- .850	.764	.66315
X PS M	.110	.368	.81684
X PL M	- .987	4.338	.10671
Y AS M	1.542	2.207	.31536
Y AL M	1.058	1.273	.50934
Y PS M	1.672	2.548	.26491
Y PL M	1.178	3.712	.14658
Z AS M	.821	1.065	.56725
Z AL M	- .337	1.914	.36579
Z PS M	.804	1.799	.38865
Z PL M	- .414	1.453	.46414
P 21 M	- .080	.889	.62129
P 22 M	.306	.282	.85505
P 23 M	.454	1.468	.46073
P 24 M	- .493	.554	.74043
P 25 M	.436	.507	.75870
Joint Test of Fit (d.f. 48)		48.030	.47163

TABLE C III

CHI-SQUARE VALUES FOR OBSERVED AND EXPECTED PROPORTIONS OF 1977 GRADER AWARDS TO EACH GRADE-SCORE FOR ESSAYS ON LOGIC GRADING DIMENSION

Lenient Graders' Mean Logit	.050
Medium Graders' Mean Logit	.636
Harsh Graders' Mean Logit	1.113

ESSAY CODE	ABILITY LOGIT ESTIMATE	χ^2 (d.f. 1.92)	p <
V AS L	- .747	1.597	.43120
V AL L	-1.032	1.544	.44293
V PS L	- .501	.426	.79217
V PL L	- .993	4.155	.11703
W AS L	- .016	5.013	.07581
W AL L	-1.196	1.142	.54515
W PS L	- .086	.392	.80656
W PL L	-1.196	1.081	.56253
X AS L	.007	.934	.60704
X AL L	- .590	1.953	.35906
X PS L	- .016	1.220	.52361
X PL L	- .473	1.186	.53271
Y AS L	1.143	1.556	.44029
Y AL L	.859	1.865	.37560
Y PS L	1.072	4.053	.12326
Y PL L	1.002	.980	.59286
Z AS L	.714	2.109	.33152
Z AL L	- .158	1.187	.53245
Z PS L	.546	.361	.81987
Z PL L	- .390	1.224	.52255
P 21 L	.329	.459	.77831
P 22 L	.097	.511	.75726
P 23 L	.787	2.368	.29050
P 24 L	.030	2.581	.26054
P 25 L	.805	.974	.59462
Joint Test of Fit (d.f. 48)		40.870	.75747

APPENDIX D

COMPLETE STATISTICAL RESULTS OF ANALYSIS OF VARIANCE OF 1976 DATA

TABLE D I

MEANS, VARIANCES AND STANDARD DEVIATIONS OF
1976 VARIABLES

VARIABLE		MEAN	VARIANCE	STANDARD DEVIATION
AS	O:	48.2	45.7	6.7602
	M:	47.0	61.5	7.8422
	L:	50.2	22.7	4.7645
AL	O:	53.6	50.8	7.1274
	M:	52.0	92.0	9.5917
	L:	53.0	23.0	4.7958
PS	O:	47.4	44.8	6.6933
	M:	45.4	40.3	6.3482
	L:	49.4	26.8	5.1769
PL	O:	54.8	84.2	9.1761
	M:	51.6	64.3	8.0187
	L:	54.2	32.2	5.6745
PL - AL	O:	1.2	11.7	3.4205
	M:	- 0.4	24.3	4.9295
	L:	1.2	3.2	1.7889
AL - PS	O:	6.2	10.7	3.2711
	M:	6.6	34.3	5.8566
	L:	3.6	4.3	2.0736
PS - AS	O:	- 0.8	10.7	3.2711
	M:	- 1.6	19.8	4.4497
	L:	- 0.8	10.7	3.2711
A - P	O:	- 0.4	39.8	6.3087
	M:	2.0	66.5	8.[illegible]8
	L:	- 0.4	22.3	4.7223
S - L	O:	-12.8	17.2	4.1473
	M:	-11.2	19.7	4.4385
	L:	- 7.6	6.3	2.5100
AP.SL	O:	2.0	5.0	2.2361
	M:	1.2	21.7	4.6583
	L:	2.0	5.5	2.3452

TABLE D II

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES PL - AL (1976)

F Ratio for Multivariate Test of Equality of Mean Vectors = 6.3760
(d.f. = 3 and 2; p<.1386)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
PL - AL O:	7.2	.6154	.4767	7.2000 / 11.7000	.6154	.4767
PL - AL M:	.8	.0329	.8649	10.6011 / 17.7194	.5983	.4956
PL - AL L:	7.2	2.2500	.2081	2.6663 / .2010	13.2664	.0679

(d.f. hypothesis = 1; d.f. error = 4)

TABLE D III

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES AL - PS (1976)

F Ratio for Multivariate Test of Equality of Mean Vectors = 4.0707
(d.f. = 3 and 2; p<.2035)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
AL - PS O:	192.2	17.9626	.0133	192.2000 / 10.7000	17.9626	.0133
AL - PS M:	217.8	6.3499	.0654	9.1137 / 10.3536	.8802	.4174
AL - PS L:	64.8	15.0698	.0179	.0013 / 1.0399	.0012	.9751

(d.f. hypothesis = 1; d.f. error = 4)

TABLE D IV

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES (PS-AS) (1976)

F Ratio for Multivariate Test of Equality of Mean Vectors = .1637
(d.f. = 3 and 2; p<.9125)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
PS - AS O:	3.2	.2991	.6136	3.2000 / 10.7000	.2991	.6136
PS - AS M:	12.8	.6465	.4665	2.8671 / 11.5950	.2473	.6532
PS - AS L:	3.2	.2991	.6136	.1608 / 1.1378	.1413	.7431

(d.f. hypothesis = 1; d.f. error = 4)

TABLE D V

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES A - P (1976)

F Ratio for Multivariate Test of Equality of Mean Vectors = 1.1927
(d.f. = 3 and 2; p<.4863)

VARIABLE		MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
A - P	O:	.8	.0201	.8942	.8000 / 39.8000	.0201	.8942
A - P	M:	20.0	.3008	.6126	29.2470 / 28.8656	1.0132	.3883
A - P	L:	.8	.0359	.8590	1.9419 / .9036	2.1490	.2804

(d.f. hypothesis = 1; d.f. error = 4)

TABLE D VI

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES S - L (1976)

F Ratio for Multivariate Test of Equality of Mean Vectors = 10.4617
(d.f. = 3 and 2; p<.0886)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
S - L O:	819.2	47.6279	.0024	819.2000 / 17.2000	47.6279	.0024
S - L M:	627.2	31.8376	.0049	5.3635 / 18.4370	.2909	.6271
S - L L:	288.8	45.8413	.0025	1.9384 / 5.4155	.3579	.6104

(d.f. hypothesis = 1; d.f. error = 4)

TABLE D VII

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES AP.SL (1976)

F Ratio for Multivariate Test of Equality of Mean Vectors = .8543
(d.f. = 3 and 2; p<.5791)

VARIABLE		MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
AP.SL	O:	20.0	4.0	.1162	20.0000 / 5.0000	4.0	.1162
AP.SL	M:	7.2	.3318	.5955	.2250 / 27.5833	.0082	.9338
AP.SL	L:	20.0	3.6364	.1293	1.7668 / 6.4185	.2753	.6522

(d.f. hypothesis = 1; d.f. error = 4)

APPENDIX E

COMPLETE STATISTICAL RESULTS OF ANALYSIS OF VARIANCE OF 1977 DATA

TABLE EI

MEANS, VARIANCES AND STANDARD DEVIATIONS OF
1977 VARIABLES

VARIABLE		MEAN	VARIANCE	STANDARD DEVIATION
AS	O:	53.4	17.3	4.1593
	M:	51.8	16.7	4.0866
	L:	54.6	12.8	3.5777
AL	O:	58.2	26.7	5.1672
	M:	57.0	23.5	4.8477
	L:	57.8	19.7	4.4385
PS	O:	53.8	13.7	3.7014
	M:	52.0	16.0	4.0000
	L:	54.6	9.8	3.1305
PL	O:	58.0	25.0	5.0000
	M:	57.4	30.8	5.5498
	L:	57.8	22.2	4.7117
PL - AL	O:	- 0.2	0.2	0.4472
	M:	0.4	0.8	0.8944
	L:	0.0	1.5	1.2247
AL - PS	O:	4.4	5.3	2.3022
	M:	5.0	1.5	1.2247
	L:	3.2	3.2	1.7889
PS - AS	O:	0.4	1.3	1.1402
	M:	0.2	0.7	0.8367
	L:	0.0	0.5	0.7071
A - P	O:	- 0.2	2.2	1.4832
	M:	- 0.6	1.3	1.1402
	L:	0.0	3.0	1.7321
S - L	O:	- 9.0	26.0	5.0990
	M:	-10.6	11.3	3.3615
	L:	- 6.4	17.8	4.2190
AP.SL	O:	- 0.6	0.8	0.8944
	M:	0.2	1.7	1.3038
	L:	0.0	1.0	1.0000

TABLE E II

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES PL - AL (1977)

F Ratio for Multivariate Test of Equality of Mean Vectors = .2667
(d.f. = 3 and 2; p<.8473)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
PL - AL O:	.2	1.0000	.3740	.2000 / .2000	1.0000	.3740
PL - AL M:	.8	1.0000	.3740	.2500 / .9167	.2727	.6377
PL - AL L:	0.0	0.0000	1.0000	.1212 / 2.2727	.0533	.8389

(d.f. hypothesis = 1; d.f. error = 4)

TABLE E III

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES AL - PS (1977)

F Ratio for Multivariate Test of Equality of Mean Vectors = 150.0667
(d.f. = 3 and 2; $p<.0067$)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
AL - PS O:	96.8	18.2642	.0130	96.8000 / 5.3000	18.2642	.0130
AL - PS M:	125.0	83.3333	.0008	6.6313 / .0975	68.0241	.0038
AL - PS L:	51.2	16.0000	.0162	1.1545 / .8065	1.4316	.3542

(d.f. hypothesis = 1; d.f. error = 4)

TABLE E IV

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES PS - AS (1977)

F Ratio for Multivariate Test of Equality of Mean Vectors = .1146
(d.f. = 3 and 2; p<.9439)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
PS - AS O:	.8	.6154	.4767	.8000 / 1.3000	.6154	.4767
PS - AS M:	.2	.2857	.6214	0.0000 / .5000	0.0000	1.0000
PS - AS L:	0.0	0.0000	1.0000	.0256 / .8205	.0313	.8760

(d.f. hypothesis = 1; d.f. error = 4)

TABLE E V

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES A - P (1977)

F Ratio for Multivariate Test of Equality of Mean Vectors = .2578
(d.f. = 3 and 2; p<.8528)

VARIABLE		MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
A - P	O:	.2	.0909	.7781	.2000 / 2.2000	.0909	.7781
A - P	M:	1.8	1.3846	.3046	1.7071 / 1.7273	.9883	.3935
A - P	L:	0.0	0.0000	1.0000	.2218 / 5.5746	.0398	.8604

(d.f. hypothesis = 1; d.f. error = 4)

TABLE E VI

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES S - L (1977)

F Ratio for Multivariate Test of Equality of Mean Vectors = 15.8289
(d.f. = 3 and 2; p<.0601)

VARIABLE		MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
S - L	O:	405.0	15.5769	.0169	405.0000 / 26.0000	15.5769	.0169
S - L	M:	561.8	49.7168	.0022	31.8181 / 4.2846	7.4261	.0723
S - L	L:	204.8	11.5056	.0275	.7172 / .7886	.9094	.4410

(d.f. hypothesis = 1; d.f. error = 4)

TABLE E VII

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES AP.SL (1977)

F Ratio for Multivariate Test of Equality of Mean Vectors = 3.8889
(d.f. = 3 and 2; p<.2113)

VARIABLE		MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
AP.SL	O:	1.8	2.2500	.2081	1.8000 / .8000	2.2500	.2081
AP.SL	M:	.2	.1176	.7489	2.4500 / .9167	2.6727	.2006
AP.SL	L:	0.0	0.0000	1.0000	.8951 / .3409	2.6256	.2466

(d.f. hypothesis = 1; d.f. error = 4)

APPENDIX F

COMPLETE STATISTICAL RESULTS OF ANALYSIS OF VARIANCE OF 1978 DATA

TABLE F I

MEANS, VARIANCES AND STANDARD DEVIATIONS OF
1978 VARIABLES

VARIABLE		MEAN	VARIANCE	STANDARD DEVIATION
AS	O:	53.6	40.3	6.3482
	M:	53.2	37.7	6.1400
	L:	55.0	36.5	6.0415
AL	O:	58.6	50.3	7.0922
	M:	56.8	36.7	6.0581
	L:	57.4	27.8	5.2726
PS	O:	53.2	27.7	5.2631
	M:	52.0	24.0	4.8990
	L:	54.4	31.8	5.6391
PL	O:	57.2	35.7	5.9749
	M:	56.4	27.8	5.2726
	L:	56.2	20.2	4.4944
PL - AL	O:	- 1.4	4.3	2.0736
	M:	- 0.4	5.8	2.4083
	L:	- 1.2	1.7	1.3038
AL - PS	O:	5.4	9.8	3.1305
	M:	4.8	3.7	1.9235
	L:	3.0	8.5	2.9155
PS - AS	O:	- 0.4	2.8	1.6733
	M:	- 1.2	2.2	1.4832
	L:	- 0.6	3.8	1.9494
A - P	O:	1.8	7.7	2.7749
	M:	1.6	12.3	3.5071
	L:	1.8	7.7	2.7749
S - L	O:	- 9.0	28.5	5.3385
	M:	- 8.0	10.5	3.2404
	L:	- 4.2	25.7	5.0695
AP.SL	O:	- 1.0	6.5	2.5495
	M:	0.8	3.7	1.9235
	L:	- 0.6	3.3	1.8166

TABLE F II

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES PL - AL (1978)

F Ratio for Multivariate Test of Equality of Mean Vectors = .8087
(d.f. = 3 and 2; p<.5942)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
PL - AL O:	9.8	2.2791	.2057	9.8000 / 4.3000	2.2791	.2057
PL - AL M:	.8	.1379	.7292	2.2326 / 3.2558	.6857	.4684
PL - AL L:	7.2	4.2353	.1088	.0429 / .1452	.2951	.6415

(d.f. hypothesis = 1; d.f. error = 4)

TABLE F III

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES AL - PS (1978)

F Ratio for Multivariate Test of Equality of Mean Vectors = 5.3074
(d.f. = 3 and 2; p<.1627)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
AL - PS O:	145.8	14.8776	.0182	145.8000 / 9.8000	14.8776	.0182
AL - PS M:	115.2	31.1351	.0051	8.4033 / 3.1701	2.6508	.2020
AL - PS L:	45.0	5.2941	.0829	.1242 / 7.7146	.0161	.9107

(d.f. hypothesis = 1; d.f. error = 4)

TABLE F IV

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES PS - AS (1978)

F Ratio for Multivariate Test of Equality of Mean Vectors = .9175
(d.f. = 3 and 2; p<.5593)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
PS - AS O:	.8	.2857	.6214	.8000 / 2.8000	.2857	.6214
PS - AS M:	7.2	3.2727	.1448	5.7202 / 2.7321	2.0937	.2438
PS - AS L:	1.8	.4737	.5292	1.1047 / 1.8039	.6124	.5159

(d.f. hypothesis = 1; d.f. error = 4)

TABLE F V

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES A - P (1978)

F Ratio for Multivariate Test of Equality of Mean Vectors = 1.8537
(d.f. = 3 and 2; p<.3693)

VARIABLE		MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
A - P	O:	16.2	2.1039	.2206	16.2000 / 7.7000	2.1039	.2206
A - P	M:	12.8	1.0407	.3654	.5141 /11.8074	.0435	.8481
A - P	L:	16.2	2.1039	.2206	3.5115 / 1.2176	2.8840	.2316

(d.f. hypothesis = 1; d.f. error = 4)

TABLE F VI

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES S - L (1978)

F Ratio for Multivariate Test of Equality of Mean Vectors = 19.9253
(d.f. = 3 and 2; p<.0482)

VARIABLE	MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
S - L O:	405.0	14.2105	.0197	405.0000 / 28.5000	14.2105	.0197
S - L M:	320.0	30.4762	.0053	8.0689 / .8743	9.2293	.0560
S - L L:	88.2	3.4319	.1376	20.9908 / 15.7980	1.3287	.3683

(d.f. hypothesis = 1; d.f. error = 4)

TABLE F VII

MULTIVARIATE ANALYSIS OF VARIANCE SUMMARY TABLE FOR
DEPENDENT VARIABLES AP.SL (1978)

F Ratio for Multivariate Test of Equality of Mean Vectors = 1.3393
(d.f. = 3 and 2; p<.4545)

VARIABLE		MEAN SQUARE	UNIVARIATE F	p<	STEPDOWN MEAN SQUARES	STEPDOWN F	p<
AP.SL	O:	5.0	.7692	.4300	5.0000 / 6.5000	.7692	.4300
AP.SL	M:	3.2	.8649	.4051	7.9506 / 2.0487	3.8808	.1435
AP.SL	L:	1.8	.5455	.5012	.1677 / .8360	.2006	.6981

(d.f. hypothesis = 1; d.f. error = 4)